Ghent University

Faculty of Medicine and Health Sciences

Department of anatomy and embryology

Department of Radiation Oncology and Experimental Cancer Research

Brachial plexus delineation in intensity modulated radiotherapy treatment planning

Joris Van de Velde

Doctoral thesis to obtain the degree of Doctor in Health Sciences

2016

Promotor: Prof. Dr. Tom Van Hoof

Co-promotor: Prof. Dr. Wilfried De Neve

Promotors:

- Prof. Dr. Tom Van Hoof
- Prof. Dr. Wilfried De Neve

Examination board:

Chair: Prof. Dr. Jan Victor

- Prof. Dr. Dominique Adriaens
- Prof. Dr. Carlo Martinoli
- Prof. Dr. Martine De Muynck
- Dr. Liv Veldeman
- Dr. Tom Verstraeten

This research was funded by the Foundation against Cancer, grant number 2012-200.

Table of contents

8.2 Developed contouring guidelines for manual BP contouring

8.3 Terms and definitions

List of abbreviations

OAR	organ at risk
IMRT	Intensity modulated radiotherapy
PTV	planning target volumes
CT	computed tomography
MRI	magnetic resonance imaging
BP	brachial plexus
C4	root between 3th and 4th cervical vertebra
C5	root between 4th and 5th cervical vertebra
C6	root between 5th and 6th cervical vertebra
C7	root between 6th and 7th cervical vertebra
C8	root between 7th cervical vertebra and 1st thoracic vertebra
T1	root between 1st and 2nd thoracic vertebra
T2	root between 2st and 3nd thoracic vertebra
RTOG	Radiation Therapy Oncology Group
RIBP	radiation-induced brachial plexopathy
RIF	radiation-induced fibrosis
Gy	Gray
PET	positron emission tomography
STIR	short-term inversion recovery
SPACE	Sampling Perfection with Application optimized Contrasts using different flip angle Evolution
3T	three tesla

DTI	diffusion tensor imaging
VIBE	volumetric interpolated breath-hold
DICOM	digital imaging and communications in medicine
CERR	computerized environment for radiation research
JI	jaccard similarity index
TAV	total agreement volume
UV	union volume
MAB	multi-atlas-based
DVF	deformation vector field
STAPLE	simultaneous truth and performance level estimation
DSC	Dice similarity coefficient
INI	inclusion index
PCC	Pearson's correlation coefficients
LSP	Lumbosacral plexus
RILSP	Radiation-induced lumbosacral plexopathy

Dankwoord

Zonder de hulp van een groot aantal mensen zou dit proefschrift er simpelweg nooit zijn gekomen. Multidisciplinaire inbreng was namelijk een absolute noodzaak voor het slagen van dit project. De mensen die onmisbaar waren in dit onderzoek wou ik allereerst nadrukkelijk bedanken.

Prof. Tom Van Hoof was de drijvende kracht achter dit onderzoek. Hij overtuigde mij om in dit onderzoek te stappen, plaveide de wegen voor mij, sprak zijn vele contacten aan en faciliteerde op alle vlak het onderzoek. Met alle problemen kon ik altijd bij hem terecht, zowel op het vlak van onderzoek als privé en meestal slaagde hij erin alles met veel humor te relativeren. De voorbije jaren waarin ik met hem mocht samenwerken leerde hij me niet enkel veel op wetenschappelijk vlak, maar nog meer over wat werkelijk belangrijk was in het leven en hoe ik alles in een juist perspectief moest plaatsen.

Prof. Wilfried De Neve was de tweede onmisbare steunpilaar van het onderzoek. Als we er echt niet uit geraakten en volledig ten einde raad waren, was er meestal slechts 1 man die ons nog kon helpen: Wilfried. Door zijn enorme ervaring en wijsheid wist hij steeds de geschikte oplossing te bedenken en door zijn inzicht stuurde hij ons steeds de juiste richting uit. Hij zorgde niet alleen voor funding, maar was voor mij altijd een onevenaarbare motivator door mij zoveel vertrouwen te schenken.

Verder wil ik **Johan Wouters** bedanken. Johan kwam via zijn masterthesis als geneeskundestudent bij ons terecht maar is natuurlijk zoveel meer dan een geneeskundestudent. Als afgestudeerd burgerlijk ingenieur met ongekende programmeursskills was hij eveneens onmisbaar voor het onderzoek. Zonder Johan was er van het technische luik van het onderzoek niets in huis gekomen. Hij vulde perfect aan de vele vlakken waar ik tekortschoot. Maar niet enkel de technische kwaliteiten van Johan zijn uitzonderlijk, ook als mens heb ik hem als zeer aangenaam en intelligent ervaren.

Nog 2 zeer belangrijke personen voor het onderzoek, bij wie ik met technische problemen of voor een aangename babbel steeds terecht kon, waren **Tom Vercauteren** en **Werner De Gersem**. Ook met mijn vele vragen over radiotherapie kon ik steeds bij hen terecht. Ze maakten steeds tijd vrij en waren altijd in voor een grapje. Oprechte dank hiervoor Tom en Werner.

Van de cel biostatistiek wil ik **Roos Colman** bedanken voor haar deskundig statistisch advies.

Verder, van de dienst radiotherapie gaat mijn bijzondere dank naar **Liv Veldeman** en **Chris Monten**: aangename dames van wie ik steeds veel steun heb gekregen en met vragen terechtkon, **Bruno Speleers**: belangrijk geweest om logistieke problemen op te lossen. Ook dank aan **Piet Ost, Thomas Mulliez, Frédéric Duprez, Frank Vanpachtenbeke, Philippe Vuye, Wendy Schelfhout** en **Katrien Vandecasteele** voor de steun aan het onderzoek.

Verder wil ik de mensen van de dienst anatomie, en in het bijzonder **Prof. Katharina D'Herde**, bedanken om me de kans te geven dit doctoraat te verwezenlijken en voor de flexibiliteit die ze daarbij aan de dag legden. **Ingrid, Marc, Liesbeth, Michael, Femke, Katrijn, Barbara, Stefanie en Dominique**: dank u voor de steun. Bijzondere dank aan **Joris Leijnse** voor de belangrijke bijdragen aan de illustraties en tekst.

Speciale dank ook aan **Emmanuel Audenaert** van de dienst orthopedie.

Van de dienst radiologie wil ik **Prof. Eric Achten** bedanken. Verder ook **Stefanie Bogaert, Pieter Vandemaele** en **Wouter Huysse**, voor het -soms avondlijke- uittesten van de vele MRI sequenties.

I would like to thank the **members of the examination committee** for their valuable comments and suggestions.

Als laatste wil ik zeer uitdrukkelijk de **Stichting Tegen Kanker** bedanken voor de financiële steun die ze me gaven om 3 jaar aan onderzoek te kunnen doen.

Ten slotte wil ik dit werk opdragen aan mijn **familie** en in het bijzonder mijn gezin: **Ellen, Elias** en **Kaatje**. Jullie zijn absoluut de 3 belangrijkste personen in mijn leven. Dank u voor de stabiliteit, rust en liefde die jullie me altijd hebben gegeven.

Joris Van de Velde
Gent, 6 juli 2015

1. General introduction

1. General introduction

1.1 Background Intensity Modulated Radiotherapy

Radiation therapy is an integral part of the treatment of patients inflicted with cancer. It is estimated that over 60 percent of patients with cancer will have radiotherapy as part of their total course of treatment (1). Over recent years, new techniques have been developed to obtain better tumor control and more effective sparing of organs at risk (OARs). The introduction of intensity modulated radiotherapy (IMRT) in the early 90's meant a significant improvement compared to the use of conventional external beam radiotherapy.

In IMRT the possibility was created to generate improved dose distributions that could be tailored to fit a complex geometric arrangement of targets that push against or even surround healthy critical structures.

This new treatment modality goes beyond the capabilities of the earlier technology called three-dimensional (3D) radiation therapy. IMRT took the older approach of using fields that conformed to the silhouette of the target to deliver a relatively homogeneous intensity of radiation and separated the conformal fields into many subfields so that intensity could be varied to better control the final dose distribution. In this way, these intensity modulated beams offer the possibility to even irradiate concave planning target volumes (PTV's) in a millimeter precise way and to avoid and control the dose delivered to organs at risk so that they are not seriously damaged in the process of irradiating nearby targets to an appropriately high dose (2).

In its current form, IMRT planning starts with line drawings of target volumes and OARs on planning-computed tomography (CT) images. The delineation may be a result of co-registration of different imaging modalities. The resulting contours are superimposed on anatomical CT image information for radiotherapy planning. CT-imaging remains the cornerstone of radiotherapy planning because it provides resolution in the order of mm, image integrity (i.e. negligible image distortion) and electron density information for accurate dose computations at reasonable cost.

Planning aims for the volumes enclosed by the CT contours are usually set as dose objectives for the target (tumor) volumes and dose constraints to the OARs. The treatment planning algorithm uses the 3D map of dosimetric goals and constraints, called dose objectives for planning, to determine a machine instruction file. Ideally the machine instruction file allows the radiation equipment to deliver exactly the desired doses. Dose prescriptions are usually written in a clinical protocol, which also specifies the patient cohort by means of patient selection criteria. It expresses the desired dose

distribution in a patient or patient cohort. On the basis of patient-specific CT imaging, impossible, irrelevant, or conflicting dose objectives may exist in the prescription. Planning systems differ substantially in their ability to deal with those problematic dose objectives.

1.2 Contouring of target and organs at risk

Since the IMRT planning process starts with the contouring of targets and OARs on planning-CT in order to determine the dose distribution, we can state that to reach their full potential, IMRT relies on the accuracy and reliability of the anatomy contouring (3).

Accurate and reliable delineations are needed to prevent an under-dosage, resulting in a decrease in tumor control probability, or over-dosage, resulting in an increase in normal tissue complication probability (3). Moreover, inaccurate delineation is one of the potential causes for uncertainties in historical dose and volume data and therefore reduced performance of predictive models (5). Unreliability of the delineations in turn is reflected in a large inter- and intra-observer contouring variability which contributes to uncertainties in radiation treatment planning which were proven to be even larger than inaccuracies due to positioning of the patient (4, 6).

Despite the importance of accurate and reliable delineations, they are tedious to achieve for multiple organs in contemporary IMRT practice.

The first potential reason for inaccurate and unreliable anatomy contouring is the inferior planning-CT quality. Image resolution and contrast of the IMRT planning-CT is often insufficient to sharply visualize most organs. Especially soft tissues (e. g. nervous structures) are hardly visible, but even details of bony landmarks are difficult to distinguish. Moreover, since in daily practice slice increment of the planning-CT usually is 3 to 5 mm, a large amount of anatomical information is lost between slices. Although other image modalities, like Magnetic Resonance Imaging (MRI), have better characteristics for the visualization of most anatomical structures, CT is chosen for radiotherapy planning above those image modalities for the previously mentioned reasons, from which the relation between the CT Hounsfield units and the electron density is decisive.

Beyond poor planning-CT quality, a second cause of inferior contouring accuracy and unreliability can be poor anatomy knowledge and insufficient delineation skills of the delineator. In an attempt to compensate the knowledge and skills of the delineator, to minimize potential delineation errors and to reduce inter- and intra-observer variability, delineation guidelines were published for every specific OAR. These delineation guidelines are written instructions by domain experts mostly illustrated with case studies (7, 9, 10, 11, 12).

However, a major problem is associated with most of these delineation guidelines. The development of the guidelines and the measurement of the accuracy and precision of image contouring by application of these guidelines is made difficult by the absence of a ground truth or 'gold standard' delineation on CT for several organs. This is particularly problematic for organs that are invisible on planning-CT, like for example nerve structures or lymph vessels. In literature, the choice of a gold standard contour for these organs varies from a mathematical average contour, an experienced radiologist-defined contour, an experienced oncologist-defined contour, or a consensus contour that is decided upon by a panel of experts. A gold standard contour that outlines the true extent of the object being contoured however is lacking. In fact, this makes it impossible to make definitive conclusions about the absolute accuracy of contours (6).

1.3 Brachial plexus as organ at risk

One of the OARs in IMRT treatment planning for head-and-neck, breast and lung cancer patients is the brachial plexus (BP). Accurate BP delineations are particularly difficult to achieve due to its poor visibility on planning-CT and the complexity of its anatomy. This large network of nerves arises from the ventral rami of spinal nerves C5 (intervertebral foramen C4-C5) to T1 (intervertebral foramen T1-T2), posterior to the vertebral artery, to innervate the upper limb.

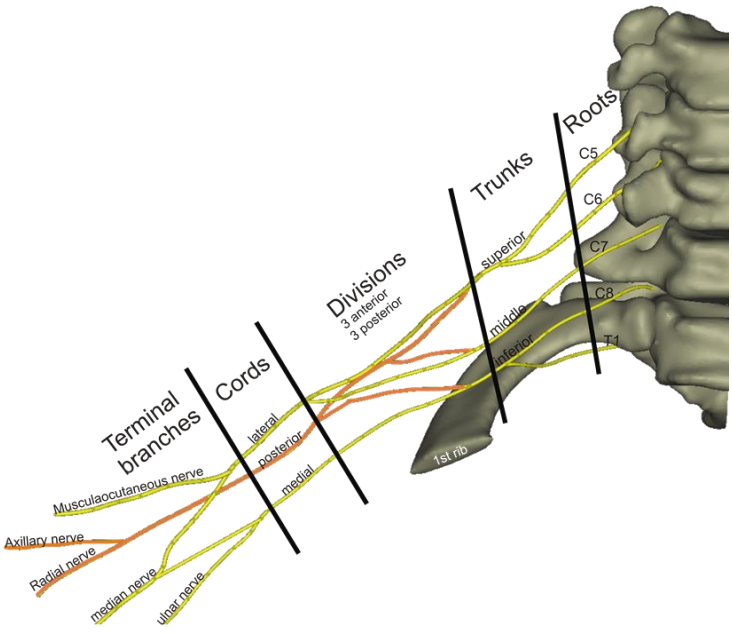


Fig. 1. Schematic illustration of the brachial plexus course.

Rami C5, C6 and C7 are situated in the grooves for the spinal nerves and ramus C8 and T1 above and beneath the neck of the first rib. These rami enter the scalene opening and fuse to form trunks. The

superior trunk is formed from the fusion of rami C5 and C6. The middle trunk is the continuation of C7 and the inferior trunk is formed by C8 and T1. Then, these three trunks enter the space between the subclavian muscle and the first head of the serratus anterior muscle and every trunk splits into an anterior and posterior division. The anterior divisions of the superior and middle trunk join to make the lateral cord, the anterior division of the inferior trunk forms the medial cord and the three posterior divisions fuse to build up the posterior cord. These cords follow their course behind the minor pectoral muscle to form the 5 terminal branches of the BP. The lateral cord forms the musculocutaneous nerve and gives the lateral contribution for the median nerve. The posterior cord splits into the radial nerve and the axillary nerve and the medial cord gives the medial contribution of the median nerve and the ulnar nerve. These five terminal branches cutaneously and motorically innervate the entire upper limb along with other smaller pre-terminal BP branches which innervate different shoulder girdle muscles. Two of these branches are purely cutaneous and innervate the medial side of the upper- and fore-arm (medial cutaneous nerve of the arm and forearm).

1.4 Brachial plexus delineation in IMRT treatment planning

Accurate BP delineation in IMRT treatment planning seems to be increasingly important due to recent developments in radiotherapy clinical practice.

First, Chen et al. (13) reported that, when no BP contouring was done and no dose constraints for the BP were established, the dose to the BP is significantly increased among patients undergoing IMRT compared with conventional radiotherapy for head-and-neck cancer. This means that with IMRT, an increased risk exists of developing BP-associated complications.

Second, there is a trend toward hypofractionation in radiotherapy to shorten the waiting time and to make the treatment more convenient for patients. However there is a concern that a short course fraction schedule may be associated with a higher risk of late complications to the BP (16).

Third, the number of long-term cancer survivors is increasing, this provides a greater opportunity for late side effects to appear, to increase in severity and to possibly have a bigger impact on the patients quality of life (14).

However, accurate BP delineations are tedious to achieve. BP delineation on planning-CT is strongly complicated by the poor visibility of the BP on CT. The location of the BP can only be deduced from the location of other CT-visible reference structures like bones or muscles. For this reason, specific delineation guidelines were developed to help clinicians contour the BP on CT.

Hall et al (7) published BP contouring guidelines for head-and-neck cancer patients that consist of 9 written points based on anatomic text books and reviewed radiological data. These guidelines are the official Radiation Therapy Oncology Group (RTOG)-endorsed guidelines.

1. Identify and contour C5, T1, and T2.

2. Identify and contour the subclavian and axillary neurovascular bundle.

3. Identify and contour anterior and middle scalene muscles from C5 to insertion onto the first rib.

4. To contour the brachial plexus OAR use a 5-mm diameter paint tool.

5. Start at the neural foramina from C5 to T1; this should extend from the lateral aspect of the spinal canal to the small space between the anterior and middle scalene muscles.

6. For CT slices, where no neural foramen is present, contour only the space between the anterior and middle scalene muscles.

7. Continue to contour the space between the anterior and middle scalene muscles; eventually the middle scalene will end in the region of the subclavian neurovascular bundle.

8. Contour the brachial plexus as the posterior aspect of the neurovascular bundle inferiorly and laterally to one to two CT slices below the clavicular head.

9. The first and second ribs serve as the medial limit of the OAR contour.

Fig. 2. RTOG-endorsed brachial plexus delineation guidelines by Hall et al.

Points 8 and 9 were later modified by Yi et al. (8).

8. contour the brachial plexus at the posterior aspect of the neurovascular bundle inferiorly and laterally one to two CT slices below the most inferior portion of the sternoclavicular joint and one to two CT slices superior to the level of the top of the aortic arch and the inferior aspect of the glenohumeral joint. The contour should have an approximate transverse length of 3 to 4 cm at this level.

9. the first and second ribs serve as the medial and approximate lateral limits of the OAR contour. The contour should not pass inferiorly to the level of the second rib.

Fig.3. Modified guidelines by Yi et al.

The second type of BP contouring guidelines for head-and-neck cancer patients were published by Truong et al (9). These guidelines are based on indirect CT and MRI images comparison of different patients and consist of 5 written steps:

1. Identify the C4–5 and T1–2 neural foramina at sagittal planning-CT to determine the upper and lower limits of the brachial plexus.

2. Contour the ventral rami of C5-T1 as they exit through the intervertebral neural foramina as seen at axial CT.

3. Contour the trunks of the brachial plexus between the anterior and middle scalene muscles.

4. Follow the insertion of the scalene muscles into the first rib.

5. Contour the brachial plexus divisions, cords, and terminal nerves by following the subclavian artery into the axilla.

Fig.4. Brachial plexus delineation guidelines by Truong et al.

1.5 Radiation-induced brachial plexopathy

The above mentioned delineation guidelines aimed to counter the inaccuracy and unreliability related to BP delineation. Inaccurate and unreliable BP contouring can cause excessive irradiation doses to the BP. When the BP tolerance dose is exceeded, patients run the risk of developing radiation-induced brachial plexopathy (RIBP).

Clinically, RIBP begins with subjective paresthesia or dysesthesia which usually decreases with the development of hypoesthesia or anesthesia. Pressure in the axillary and/or supraclavicular zone can trigger this paresthesia. Neuropathic pain is generally rare and moderate, except after failure of neurolysis. Motor weakness is progressive, often delayed by several months, and associated with fasciculations. The topography of symptoms varies with the level of plexus damage. It frequently starts at the median nerve, simulating carpal tunnel syndrome, before spreading progressively to the forearm and then the upper arm. It is caused by local damage to mature nerve tissue, which is partly attributable to initial microvascular injury, and partly to radiation-induced fibrosis (RIF). RIF is a

dynamic process related to perturbations at various levels of physiological homeostasis, varying from inflammation to sclerosis, and is characterized by gradual stepwise worsening over a period of several years (15).

The onset is often insidious, occurring over several months or years. The interval between radiotherapy treatment and the onset of RIBP depends strongly on the total dose and the dose per fraction. This interval becomes shorter when dose per fraction and/or total dose increases. (16). For this reason, the time to the appearance of symptoms reported in literature may vary significantly. Amini et al. (17) reported a median time to onset of RIBP of 6.5 months (range 1.4 - 37.4 months). Chen et al. (13) found a median time interval from completion of radiotherapy to the development of neuropathic symptoms of 33 months (range 10-72 months). Delanian et al. (15) reported a time to onset that ranges from several months to decades. The intensity of RIBP is variable, but progressively increases, and after several years may result in paralysis of the upper limb in a range of 2 months to 5 years from the first signs to hand paralysis (19).

RIBP is a radiotherapy side-effect in head-and-neck, breast and lung-top cancer patients.

RIBP was thought to be uncommon for head-and-neck cancer patients, but recent clinical investigations suggest that it may be underreported (13, 20).

Chen et al. 2011 (13) prospectively screened 330 patients who had previously completed radiotherapy for head-and-neck cancer. Twelve percent of them reported neuropathic symptoms and when patients with less than five years follow up were excluded, the rate of positive symptoms increased to 22 percent. In a more recent publication, Chen et al. (18) prospectively screened 352 head-and-neck cancer patients. Fourteen percent of them reported BP related neuropathic symptoms. Specifically, a major concern for RIBP arises when a need occurs to treat the lower neck at high doses in a post-operative or level IV node-positive neck.

The occurrence of RIBP with head-and-neck cancer treatment is consistent with RIBP observed in breast cancer patients. In breast cancer patients also the lower neck is treated in cases of node positive disease. Lundstedt et al. (22) reported a significant difference in incidence of RIBP in breast cancer patients treated with axillary dissection + radiotherapy (20%) and patients treated with axillary dissection alone (13%). A recent review from Delanian et al. (15) showed up to 12% brachial plexopathy after mastectomy + axillary dissection levels I-II when using 50 Gy in 25 fractions.

For lung cancers near the apical region, RIBP is a major concern for high-dose radiotherapy. Amini et al. (17) found that 16% of the patients treated with radiotherapy for lung cancer with a median irradiation dose of 70 Gy had RIBP.

In cancer survivors, RIBP arises often months or even years after radiotherapy treatment. This makes that the link with previous radiotherapy is forgotten or difficult to establish. The diagnosis is then made after a long series of medical consultations and tests. The diagnosis is based on neurological expertise by analysis of symptoms, electrophysiological findings, MRI and Positron Emission Tomography (PET)-scan. MRI and PET imaging may provide a further tool in ruling out suspected malignant plexopathy. The radiotherapist, in collaboration with the neurologist, can assess with a dosimetric reconstruction whether the neurological symptoms can be related to nerve damage within the irradiation volume.

The prognosis of RIBP is poor, with most patients experiencing either a rapid or insidious worsening over time (23).

Risk factors to develop RIBP are threefold. Firstly, the radiotherapy related risk factors like a large total dose or a large dose per fraction. Secondly, the combined treatment related factors like previous lymph node dissection, concomitant or previous neurotoxic chemotherapy. And thirdly, patient related factors like young or advanced age, obesity, co-morbidity factors such as high blood pressure, combined peripheral neuropathy and diabetes mellitus.

RIBP treatment is symptomatic. Pain is treated by antidepressant and antiepileptic drugs. Surgical methods are not proven to be useful in the management of RIBP. Physical therapy is valuable in maintaining function and preventing joint complications (15).

General recommendations for the dose tolerance of the BP are based on a more than twenty year old study by Emami et al (24). They found a five percent risk at five years if one-third of the plexus received 62 Gy or if two-thirds received 61 Gy or 100 percent received 60 Gy. They also reported a 50 percent risk at five years corresponding to 77 Gy, 76 Gy and 75 Gy respectively.

Most studies since then have recommended that the maximum dose should be kept under 66 Gy in 2 Gy per fraction (7, 9, 10).

However, Amini et al. (17) reported that irradiation doses of that magnitude often result in local failure, which itself can cause brachial plexopathy. Also, Lundstedt et al. (22) conclude in their study that radiotherapy to the supraclavicular lymph nodes after axillary dissection increases the occurrence of brachial plexopathy when the breast area was treated only to 50 Gy. In their most recent study (21), Chen et al. recommend to keep the volume of the BP receiving more than 70 Gy (V70) less than 10 percent.

Generally, we can conclude that the true tolerance dose for the BP is not exactly described in literature. This may be due to a variability in individual radiosensitivity and even differences in radiosensitivity of various anatomic regions of the BP itself (21).

1.6 Definition of the problem

There are several problems associated with the existing BP contouring guidelines. The basic problem with the existing RTOG-guidelines for BP delineation on CT is that they are based on gold standard delineations defined by experts, without any form of validation afterwards. In fact, this method of gold standard development is very unreliable and subjective, especially for CT-invisible and anatomically complex structures like the BP. Moreover, wrong judgments are made about the accuracy of BP contours in numerous validation studies (9, 13, 18, 21, 26) because in all these studies, the same gold standards were used for accuracy assessment.

A second problem is that the guidelines are imprecise, regardless of their above stated development method. The location of the BP is too variable in relation to the reference points used. For example, contouring guidelines for regions of the BP lateral to the interscalene triangle are still referred to as being at a fixed vertebral level. This recommendation is untenable based on the degree of variation in this region, increasing with distance from the vertebral column. This is primarily due to the highly mobile claviculoscapular complex, which provides a bony interface to determine the course of the BP. Also, additional details are not clearly defined and thus subject to interpretation. For example, should the T1-T2 intervertebral foramen be delineated? Moreover, the use of a variety of terms for anatomic orientation and the designation of the BP as a reference point can lead to ambiguity.

Moreover, the existing guidelines are incomplete. They are elaborated only for a limited number of slice levels. Therefore, when non indicated slice levels need to be delineated, key information is lacking. As a result, it is difficult for clinicians to achieve a high-quality contour delineation of the BP.

Another problem that is not particularly associated with the existing contouring guidelines, but concerns BP delineation in general, is that BP segmentation on CT is a tedious and time-consuming process, so that clinicians are reluctant to do this, despite the growing importance, due to the previously described recent changes in radiotherapy clinical practice, of imposing dose constraints to the BP to prevent RIBP and for dose reporting and plan evaluation.

1.7 Aims of doctoral thesis

This doctoral thesis aimed to develop an accurate, precise and time-efficient BP segmentation method on CT for IMRT treatment planning.

Subgoals:

1. Development of gold standards for BP delineation.
2. Validation of the existing delineation guidelines for manual BP contouring.
3. Development of new BP delineation guidelines for manual contouring
4. Development of a multi-atlas based automatic BP contouring method
    a. Determination of optimal label fusion algorithm
    b. Determination of optimal number of atlases for multi-atlas based BP contouring
    c. Development of an atlas selection procedure for multi-atlas based BP contouring

2. Materials and methods

2. Materials and methods

2.1 Development of gold standards for BP delineation on CT

2.1.1 In vivo brachial plexus imaging

For the development of accurate gold standards for BP delineation, optimal image modalities have to be chosen. Since CT-scan is not able to visualize the BP in detail, we were forced to seek refuge in other imaging modalities. MRI is the imaging modality of choice for the evaluation of the BP due to its superior soft tissue resolution and multiplanar capabilities. Most publications (27, 28, 29) cite the T2 3D STIR (Short-Tau Inversion Recovery) SPACE (Sampling Perfection with Application optimized Contrasts using different flip angle Evolution) sequence on a 3T MRI scanner as the best option for BP imaging. However, our own experience with in vivo BP imaging on MRI was that motion and respiratory artefacts, in combination with the complexity of the BP anatomy and its distribution in space, make it impossible to visualize the complete BP anatomy in a detailed way with the above mentioned sequences.

Diffusion tensor imaging (DTI) is a very promising technique but 3D reconstruction of the complete BP is still not possible with DTI. The relatively small size of the nerve fibers, the orientation, and the localization of the BP between the neck and the shoulder, resulting in considerable geometric distortion and artefacts, make DTI of the BP very challenging and only BP roots have been clearly visualized with this technique (27, 30, 31, 33).

2.1.2 In vitro brachial plexus imaging

To completely undo motion and breathing artefacts and to allow dissection for validation, cadaver specimen were used in a first stage as a baseline for BP imaging. Fifteen cadavers embalmed following the Thiel method (34) were included for BP imaging. Thiel embalmed cadavers are specifically characterized by flexible soft tissues and joints, full range of motion, and maintenance of tissue integrity. Furthermore, the use of Thiel cadavers has previously provided very accurate models, resulting in high-quality data for biomechanical and dosimetric studies (35- 38). In addition, Thiel cadavers provide the opportunity for long MRI scanning times, unlimited radiation exposure (for CT), optimal standardization of body posture, and exclusion of movement, flow, and breathing artifacts.

A disadvantage of using Thiel cadavers is the potential occurrence of diminished signal and contrast observed during imaging the cadavers. This may be attributed to the high conductivity of the embalming liquids (29).

Through a 'trial and error' procedure, several sequences and parameter settings were tested on Thiel cadavers for their capability to visualize nerve tissue embedded in adipose tissue with high contrast. A 3D T1-weighted fat suppressed Volumetric Interpolated Breath-hold (VIBE) MRI sequence (voxel 0.8 × 0.8 × 0.8 mm, 320 mm field-of-view, 384 × 384 pixel, 0.8 mm slice increment, TE= 1.48 ms, TR= 4.12 ms and a scanning time of 32 min) with a 12-channel head coil, a 4-channel neck matrix coil, a 6-channel body matrix coil and a 24-channel spine matrix coil came out to be the best for BP visualization. This spoiled gradient-echo sequence can acquire T1-weighted images with high quality and spatial resolution in a relatively short acquisition time. It was originally applied to abdominal examinations and since then has gained wide acceptance for contrast-enhanced T1-weighted abdominal imaging (38, 39). In addition, fat-suppressed 3D VIBE has also been used in brain and chest imaging, whole-body tumor staging and in articular cartilage imaging (41). To our knowledge, this MRI sequence was not previously used to visualize the BP.

2.1.3 3D Reconstruction

Cadaver MRI Digital Imaging and Communications in Medicine (DICOM)-images were imported in Mimics® and visualized in 3 planes (coronal, sagittal and axial). Mimics® image processing software (Materialise, Leuven, Belgium) is developed to create 3D surface models from stacks of 2D image data.

A brachial plexus 3D reconstruction was created using the procedure published by Van de Velde et al. (Van de Velde et al. 2015). A predefined MRI grayscale was chosen and manually adapted until the optimal contrast for BP visualization was achieved. In the axial view, a mask was created manually for the BP with the 'edit mask' tool. In every axial MRI slice, the detailed position of all the BP parts was determined and contoured. These axial 2D contours were reconstructed by the '3D calculation tool'. In this way, a detailed 3D reconstruction was made showing all the different rami, trunks, divisions, cords and terminal branches of the BP. For improved visualization, 'smoothing' was performed with compensation for shrinkage (Iterations 1, smooth factor 1).

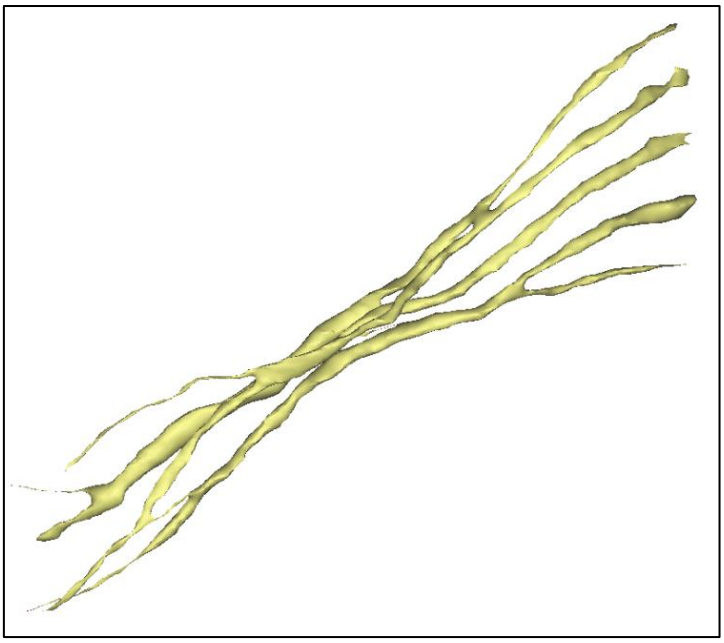


Fig. 5. Brachial plexus 3D reconstruction

2.1.4 Anatomical validation

In a next step, the virtual 3D reconstructed plexus was anatomically validated. Since every BP is unique in its branching pattern, the accuracy of the 3D reconstruction can only be validated by a cadaver dissection. The shoulder region of each cadaver was dissected according to an in-house protocol (38, 46) developed to preserve local topography. Afterwards, the virtual BPs were visually compared with the dissected BPs in the corresponding cadaver. The presence of atypical branching patterns and anatomical variations in the dissected plexuses were verified on their presence in the corresponding 3D reconstruction for extra accuracy evaluation.

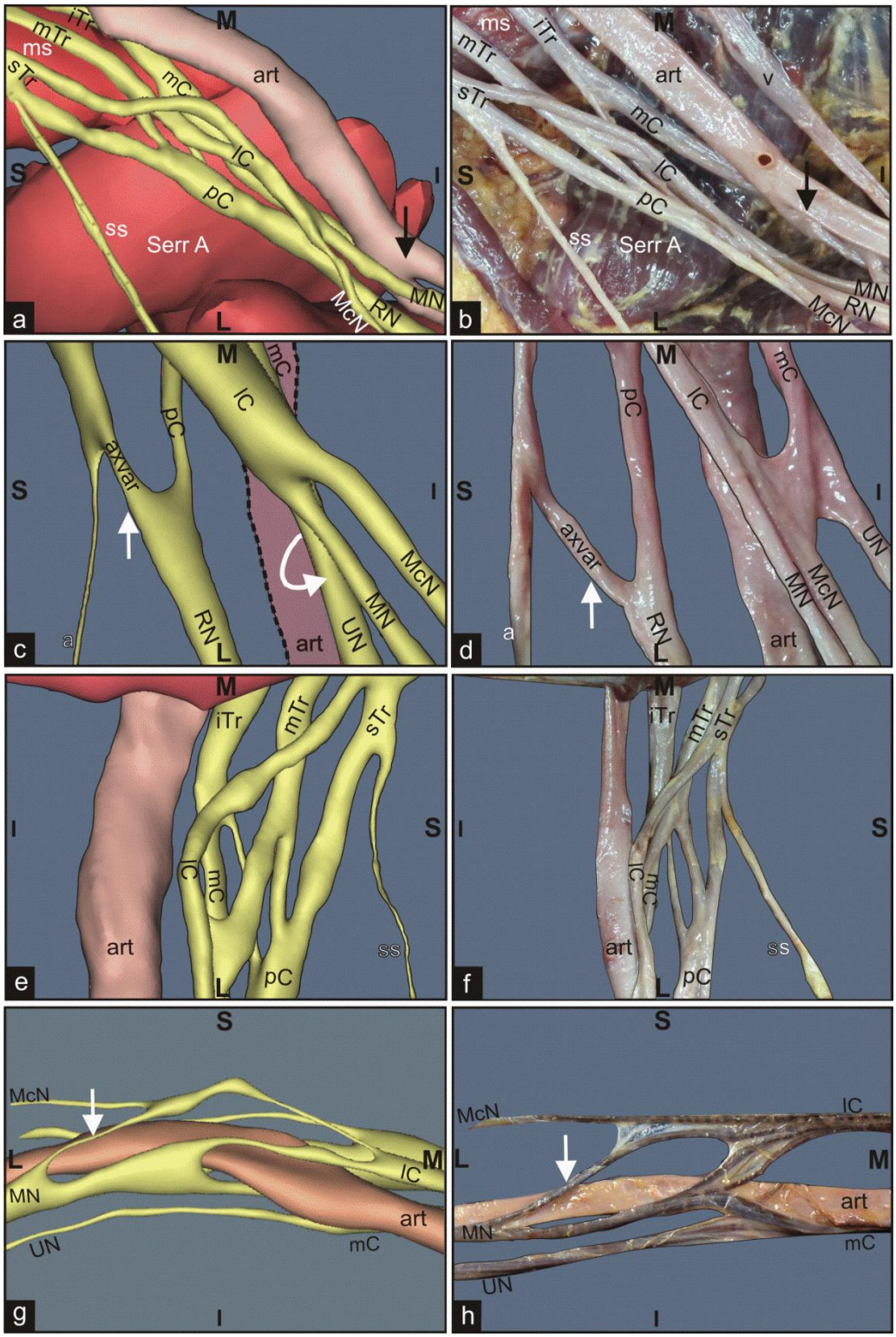


Fig. 6. 3D-reconstructions of brachial plexuses are anatomically validated by cadaver dissection. Brachial plexus details of four different cadavers are presented. In the first cadaver **(a, b),** a bifurcation of the brachial artery is visible in both reconstruction **(a)** and dissection views **(b)** (black arrows). In the second cadaver **(c, d)**, an accessory axial nerve (axvar) connected with the posterior cord is present in both reconstruction and dissection views (white arrows). The axillary artery (dotted line) was not reconstructed, the median and musculocutaneous nerves are rotated around the ulnar nerve (curved arrow, reconstruction view) due to abduction-exorotation position of the shoulder in the dissection view. In the third cadaver **(e, f)** the branching pattern of the reconstruction and the

dissection view are identical. In the fourth cadaver **(g, h)**, an accessory lateral contribution of the lateral cord to the median nerve is present in both reconstruction and dissection views (white arrows).

sTr, superior trunk; mTr, middle trunk; iTr, inferior trunk; pC, posterior cord; mC, medial cord; lC, lateral cord; ms, middle scalene; Serr A, serratus anterior; subsc, subscapular muscle; art, axillary artery; v, axillary vein; a, axillary nerve; RN, radial nerve; UN, ulnar nerve; MN, median nerve; McN, musculocutaneous nerve; ss, suprascapular nerve; M, medial; L, lateral; S, superior; I, inferior.

2.1.5 CT-MRI fusion

Since the IMRT contouring process is performed on CT and not on MRI, CT had to be obtained for each cadaver in an identical treatment position. Therefore, cadavers were positioned in a standardized IMRT treatment position for CT and MRI on a carbon plate and fixated by a head-and-neck thermoplastic mask, conform to daily practice. A helical CT scan (Toshiba Aquilion) with the following scanning parameters: 120 kVp, 300 mAs, slice increment 1 mm, 502.78 mm field of view, and 512 x 512 pixels (0.982 mm pixel size) was taken.

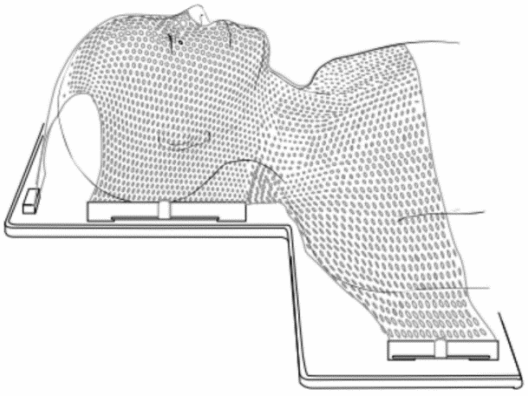


Fig. 7. Illustration of a head-and-neck thermoplastic mask

After imaging, CT and MRI images were rigidly fused by the 'image registration' function in Mimics®. The 'image registration' function allows to fuse two datasets by doing a landmark point based registration. The transformation matrix obtained by this fusion was used afterwards to correctly position the MRI BP reconstruction in the coordinate system of the corresponding CT. The result was an anatomically validated gold standard BP contour on CT.

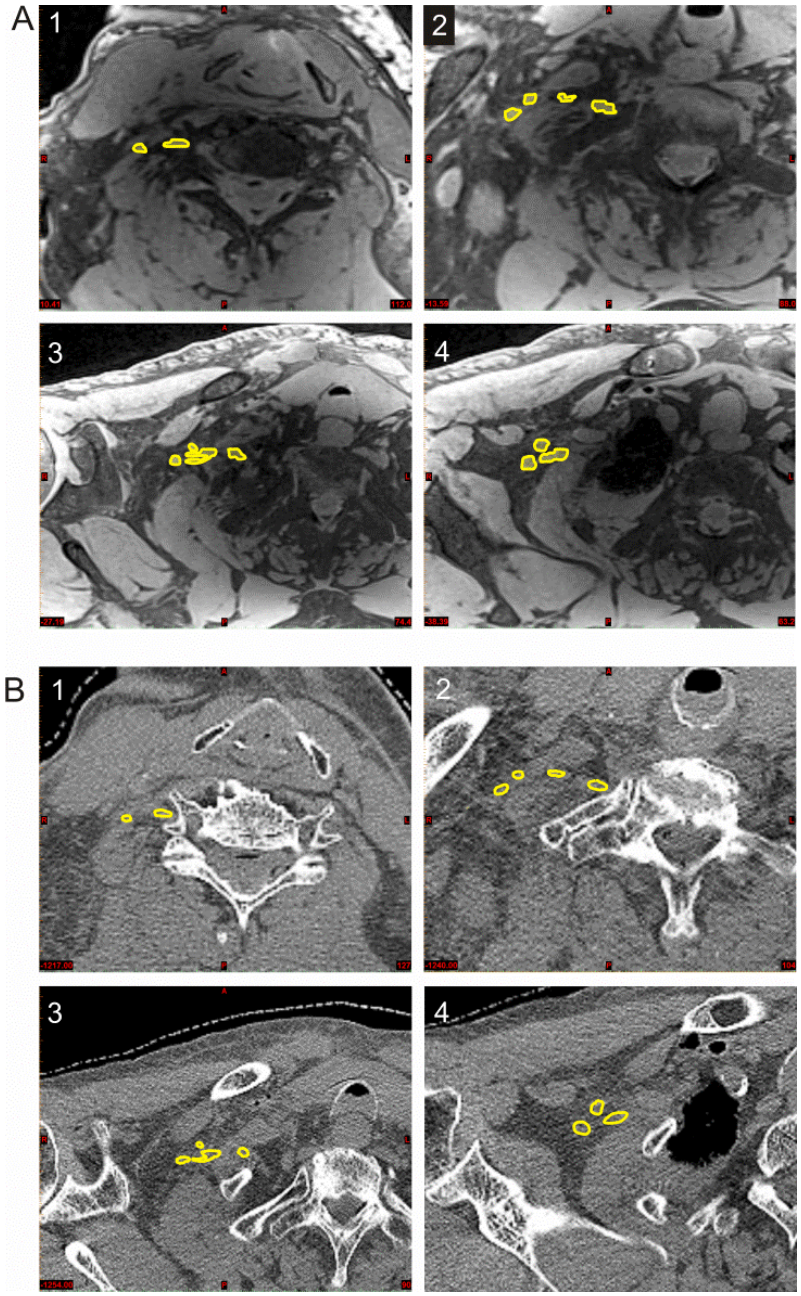


Fig. 8. The anatomically validated BP contours on MRI (A) and CT (B) in four different BP parts. 1, BP roots; 2, BP trunks; 3, BP divisions; 4, BP cords.

2.2 Validation of existing brachial plexus delineation guidelines

To adequately improve the existing BP delineation guidelines (Hall et al. 2010), their accuracy and reliability had to be investigated in detail first.

To assess inter- and intra-observer reliability and accuracy of BP contouring following the existing guidelines, five observers (two experienced radiation oncologists, three experienced dosimetrists)

participated in the BP delineation process. In Mimics®, every observer contoured the right BP on CT for the three cadavers on three different days in random order, using the RTOG-endorsed BP contouring guidelines (7). This procedure was repeated three times per cadaver for each observer with a 2 week time interval.

The BP delineations were then divided into four independent regions for detailed analysis. Region 1 extended from the origin of the BP at the intervertebral foramina until its entrance in the scalene opening. Region 2 started at the entrance of the scalene opening until the entrance between the subclavius and serratus anterior muscles. In region 3, the BP was defined between the subclavius and serratus anterior muscles. In region 4, the BP was bordered between the minor pectoral, subscapular and serratus anterior muscles.

Statistical analysis

Computerized Environment for Radiation Research (CERR) was used to analyze the reliability of the delineations. CERR is an open-source Matlab-based radiation therapy planning analysis tool. Both the inter- and intraobserver reliability were characterized using the statistical parameters described below.

In each region, contours were compared for agreement by calculating the apparent volume overlap, which is the average agreement probability by which a voxel is selected by the observers (interobserver agreement) or repeatedly by one observer (intraobserver agreement). This value was corrected for agreement by chance by using the generalized kappa statistics (42- 45).

We also calculated the Jaccard index (JI) as the ratio of the intersection volume and the entire union volume of the delineations: $JI(A,B) = (A \cap B)/(AUB)$. The JI is also situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

To measure the interobserver reliability, each observer's first delineation was imported into CERR. This was repeated for each cadaver. Afterwards the observer's second and third delineations were imported in a similar manner. The parameters were expressed, as the mean value of the three delineations for each observer, for each region separately.

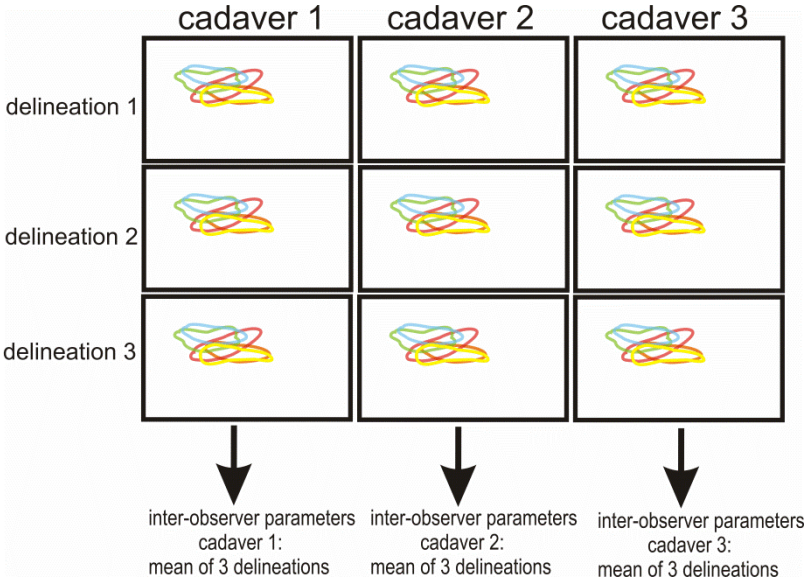


Fig. 9. Illustration of the procedure followed for the inter-reliability parameter calculation. This procedure was repeated for every delineation zone. The delineations of the 5 delineators are depicted in 5 different colors.

To measure the intraobserver reliability, the three delineations of each observer for a single cadaver were imported into CERR. This was repeated for each cadaver. To calculate intraobserver reliability, the different parameters were expressed as the mean of the three cadavers.

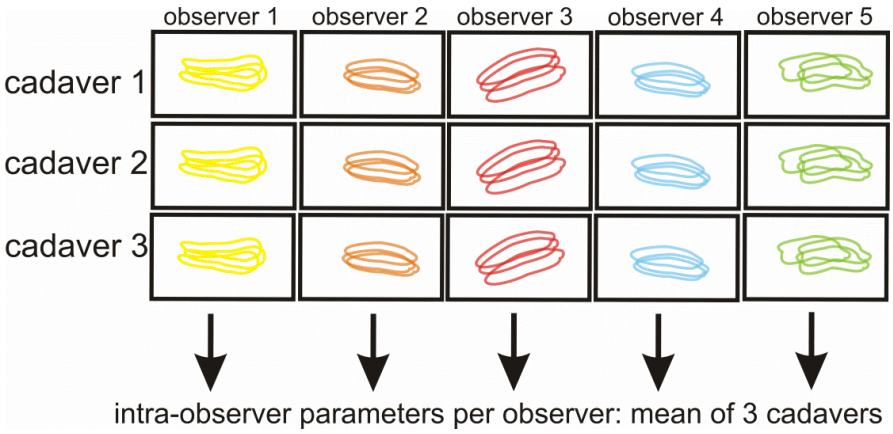


Fig. 10. Illustration of the procedure followed for the intra-reliability parameter calculation. This procedure was repeated for every delineation zone. The delineations of the 5 delineators are depicted in 5 different colors.

For accuracy assessment of the delineations, each observer's delineation was imported into the corresponding enhanced CT-project file and compared with the anatomically validated BP contour as the gold standard. Therefore, inclusion index (INI) was measured between the gold standard BP (A)

and the delineated BP (B). INI is the intersection volume of these, divided by the gold standard BP: INI = (A∩B/A). It is situated between 0 and 1 with 0 indicating no inclusion and 1 indicating the total inclusion of A by B.

2.3 Development of manual brachial plexus delineation guidelines

After validation of the existing guidelines, the BP gold standard delineations were used to develop new BP contouring guidelines.

For this purpose, BP surrounding soft tissue structures were supplementary delineated on MRI in a semiautomatic way and precisely located in the CT coordinate system by applying the transformation matrix from the initial CT-MRI image registration. This finally resulted in a CT project file with integration of the BP gold standard together with the surrounding soft tissues from the corresponding MRI.

Then, to provide an integration of anatomical variation, the individual CT project files were superimposed onto 1 'master' CT project by use of best fit principle (translation, rotation and scaling). These local coordinate transformations were defined by the respective morphology and position of proximal anatomic reference landmarks. The transformation procedure was driven by a stepwise integrated Matlab code that used a point correspondence and Procrustes function.

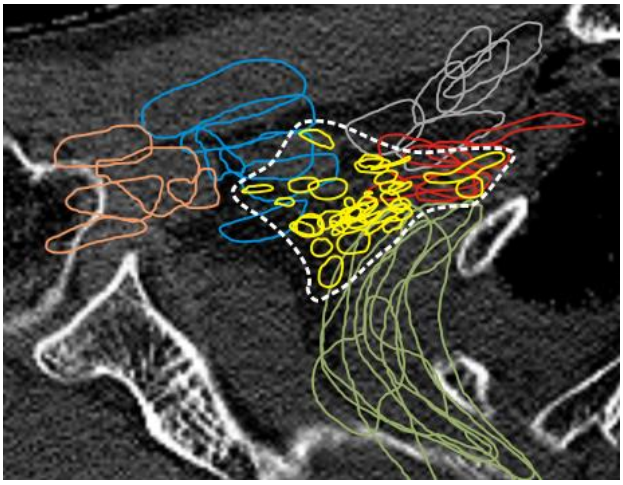


Fig. 11. 2-dimensional axial section through the 3-dimensional superimposed structures of 6 specimens for the fourth independent region. Clusters of superimposed muscles (gray, subclavian muscle; blue, minor pectoral; orange, coracobrachial muscle; green, serratus anterior) and axillary arteries (red) are visible. The BP regions are superimposed (yellow). The delineation area is identified as a dotted line.

This superimposing procedure was repeated for 4 consecutive independent regions from mediosagittal to lateral separately because the topography of the entire BP cannot be related. For example, if only the vertebral columns are superimposed as a reference, the location of the BP will

be accurate according to its medial regions, which are close to the vertebral foramina, but it will vary significantly in its lateral regions as a result of intersubject variations relative to the vertebral column. In the first independent region (vertebral region), the vertebral bodies were superimposed for each cervical level. In the second region (scalene outlet), superimposition of the scalene muscles was performed. In the third region (sublavius-serratus outlet), the subclavian, serratus anterior, anterior scalene, and minor pectoral muscles were fitted because of their close relationship with the BP. Last, in the fourth region (the muscle triangle), the axillary artery; the minor pectoral, subclavian, serratus anterior, and coracobrachial muscles; and the second rib were fitted.

After this 3D fitting procedure, the overlaid BPs and surrounding regions were visualized in axial slices as a cross-sectional clustering of superimposed contours.

Then, for each axial slice, a minimal yet convenient delineation area was established to encapsulate the superimposed BPs. The clusters of contours, corresponding to the specific reference structures, were also used to deduce the appropriate general guidelines for describing the delineation process used to achieve suitable BP inclusion for each individual case.

2.4 Development of automatic BP segmentation method

Manual BP segmentation on planning CT for radiation therapy treatment planning is a time-consuming and unreliable process (46). An effective automatic BP contouring method could relieve clinicians of this tedious task and would result in a higher inter- and intra-observer reliability and accuracy of the contouring process (47). This issue is of growing importance following the introduction of both function-sparing and adaptive IMRT, where the number and frequency of delineation of OARs are increased (48, 49).

Multi-atlas-based (MAB) autosegmentation has been demonstrated to significantly improve segmentation accuracy compared to segmentation based on a single atlas (50- 52). In MAB autosegmentation, several available presegmented images – called atlases – are first registered separately to the patient using deformable image registration. During the deformable image registration process a deformation vector field (DVF), describing the non-linear transformation from a presegmented image dataset to a patient image dataset, is created. Based on the computed DVF, a set of delineations on the presegmented image data set are deformed on the patient image data set. The multiple deformed delineations on the patient image data set are combined by the label fusion algorithm to obtain a unique and final consensus segmentation. More information about MAB autosegmentation can be found in appendix 8.3.

To develop accurate automatic MAB BP segmentations, multiple parameters have to be controlled.

First the optimal deformable image registration and label fusion algorithm have to be determined. Then, the optimal number of atlases to use has to be investigated and an effective atlas selection procedure to select the most appropriate atlases for MAB BP contouring has to be developed.
For the investigation of these different parameters related to BP autosegmentation, 15 cadaver CT-atlases with gold standard BP segmentations that were developed following the above described procedure, were used. Three cadaver CT-atlases were excluded due to the presence of metal artefacts in the cervical region.

2.4.1 Determination of optimal label fusion algorithm and number of atlases

For the determination of the optimal label fusion algorithm and number of atlases, the ADMIRE® software 1.10.02 (Elekta AB, Stockholm, Sweden) was used. ADMIRE® performs the segmentation of a novel subject image (here called 'patient') by using multiple pre-segmented images, which are also known as 'atlases'. The general algorithm in ADMIRE® is used for the initial deformable image registration (53). This atlas registration framework consists of three major steps: a linear registration and two non-linear registration steps. For label fusion, 2 different algorithms in ADMIRE® are compared: the Simultaneous Truth and Performance Level Estimation (STAPLE) label fusion (54) and Patch label fusion (55).

The STAPLE algorithm was originally designed for the validation of image segmentations. It considers a collection of segmentations and computes a probabilistic estimate of the true segmentation and a measure of the performance level represented by each segmentation (53, 54). It completely ignores the image data and uses only the segmentations when computing the label fusion. In contrast, the Patch algorithm considers the accuracy of the initial atlas registration by comparing the intensity similarity between the atlas and the patient after being aligned, to get better label fusion results. This process, is called 'intensity weighting'. More information about STAPLE and Patch label fusion can be found in addendum 8.3.

Procedure

For the determination of optimal label fusion algorithm, a leave-one-out strategy was followed. One of the 12 available cadaver CT-datasets was selected as a patient and the 11 remaining CT-datasets, which contained the anatomically validated BP segmentation, served as atlases. All of the atlases were first registered separately onto the patient using the general registration algorithm in ADMIRE®. Next, the label fusion was performed, with both STAPLE and Patch, first using every possible combination of 2 atlases. Subsequently, label fusion was repeated with a gradually increasing number of atlases, until every possible combination of 11 atlases was reached (Fig. 12). This process was reiterated for every atlas as a patient. It resulted in 24432 combinations over the

different number of atlases. A Power analysis was executed (power π= 80) to calculate the minimum sample size required for a 90% confidence interval.

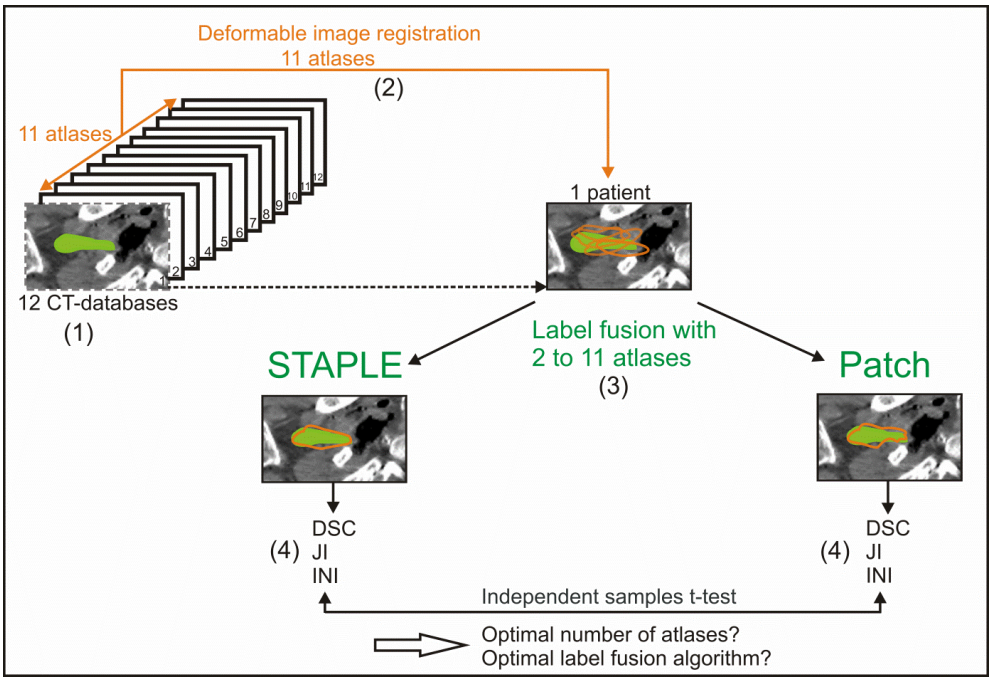


Fig. 12. Schematic illustration of the procedure for determining the optimal number of atlases and optimal label fusion. (1) 12 cadaver CT datasets were included, one atlas was selected as a patient. (2) The 11 remaining atlases were taken for deformable image registration on the patient. (3) Label fusion was done with 2 up to 11 atlases, once using STAPLE and once using Patch. (4) For each number of atlases, average Dice similarity coefficient (DSC), Jaccard index (JI) and Inclusion index (INI) were calculated for the generated contour (orange) with the gold standard contour (green). This procedure was repeated for every atlas as a patient.

Next, for every generated 'label fused' autosegmentation, 3 similarity indices with the gold standard contour were calculated to quantify their similarity:

First, the Dice similarity coefficient (DSC) was calculated between these 2 segmentations. The DSC measures the spatial overlap between the gold standard A and the registered image B, and is defined as DSC(A,B) = 2(A∩B)/(A+B) where ∩ is the intersection volume. The DSC is situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

We also calculated the Jaccard index (JI) as the ratio of the intersection volume and the entire union volume of the delineations: JI(A,B) = (A∩B)/(AUB). The JI is also situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

At last, inclusion index (INI) was measured between the gold standard BP (A) and the registered BP (B). INI is the intersection volume of these, divided by the gold standard BP: INI = (A∩B/A). INI is situated between 0 and 1 with 0 indicating no inclusion and 1 indicating the total inclusion of A by B.

Finally, for each number of atlases, average DSC, JI and INI were calculated over the different combinations. More information about the similarity indices can be found in addendum 8.3.
To determine the clinically relevant optimal number of atlases, an equivalence trial was conducted (56, 57). An equivalence trial is used to demonstrate similarity between compared groups. An equivalence margin represents a clinically acceptable range of differences. For this study, an equivalence margin of 10 percent was predetermined.
Next, the difference between STAPLE and Patch label fusion was determined using an independent sample t-test. Therefore, in the 2 label fusion groups, the similarity indices for their respective clinically relevant optimal number of atlases were compared.

2.4.2 Development of an atlas selection procedure for multi-atlas based BP contouring

To determine which atlases have to be selected to obtain the most accurate autosegmentation results, an atlas selection procedure for MAB BP contouring had to be developed. Therefore, one atlas was selected in ADMIRE® to serve as a patient, and the other 11 atlases were registered separately onto the 'patient' using deformable image registration in ADMIRE® (52). This was repeated for every single atlas as a patient. This process resulted in 11 registered brachial plexuses onto every single atlas (n= 12) (11x12= 132 total deformations).

Then, for every registered BP separately, similarity indices for the registered BP with the original gold standard BP were calculated. DSC, JI and INI were calculated between the two segmentations.

Definition of morphometric parameters

To define morphometric parameters that potentially could influence the BP segmentation quality, three points on neighboring bony structures with a stable location in relation to the BP, were indicated on every atlas. In all of the atlases, the BP generally had a triangular shape upon simplification that was defined by i) the anterior tubercle of the transverse process of C5, ii) the most anterior point of the head of the first rib and iii) the infraglenoid tubercle (the most caudal point of the glenoid fossa).

Based on these 3 reference points, several parameters were defined that could potentially influence the automatic BP segmentation.

First, absolute distances (length) between these 3 points were calculated (Fig. 13; A, B, C; mm). Second, the surface area's (A x h/ 2; mm^2) of the triangle were calculated. The third parameter was the protraction-retraction distance ('protraction'), which was defined as the horizontal distance measured in the sagittal plane, between a vertical line through the anterior tubercle of C5 and a

vertical line through the infraglenoid tubercle (Fig. 13; D). The fourth parameter was the lateralization parameter ('lateralization'), which was defined as the shoulder width/neck height ratio (Fig. 13; A/B). The final investigated parameter was elevation-depression ('elevation'). Thus, the vertical distance between the horizontal line through the anterior tubercle C5 and the horizontal line through the infraglenoid tubercle was measured (Fig. 13; E).

Finally, for every pair of atlases, the difference in morphometric parameters (Δ parameter) was calculated.

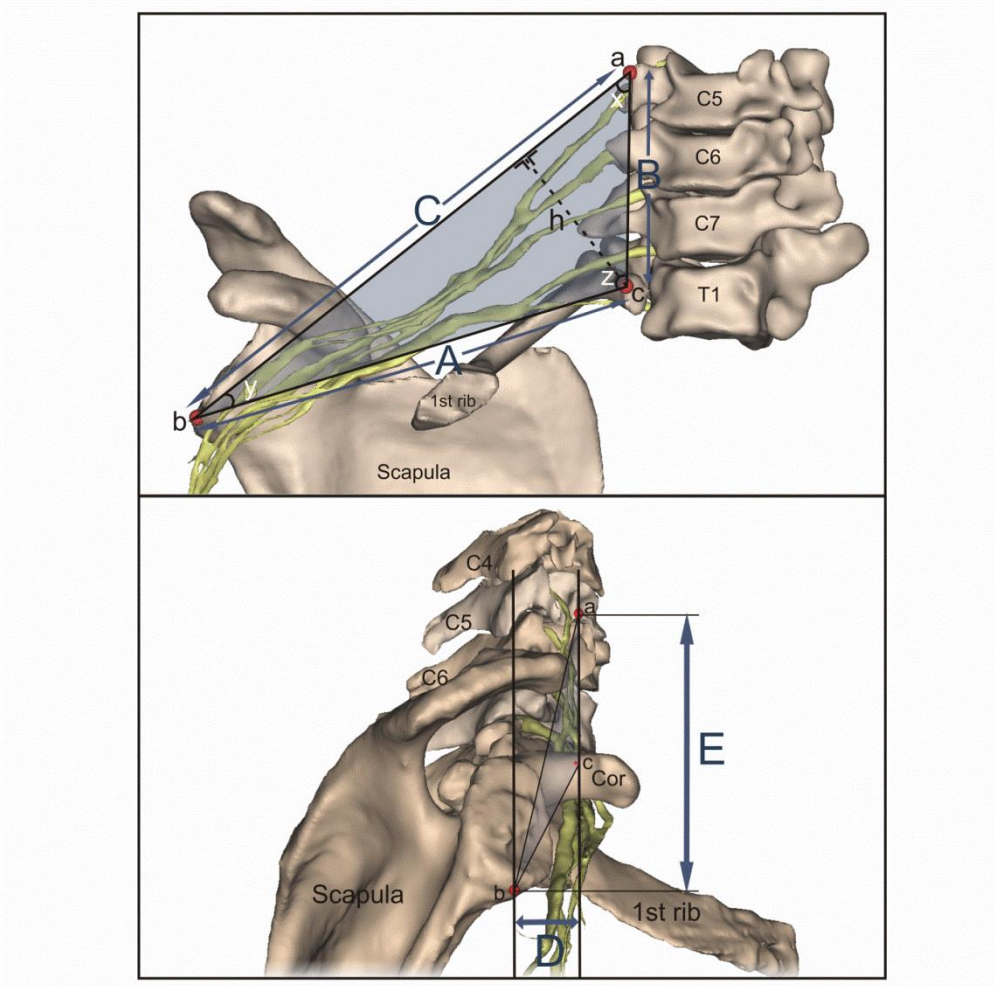


Fig. 13. Frontal (above) and lateral (below) view of a 3D illustration of the indication of the bony points nearby the brachial plexus (yellow) in a CT dataset. (a, anterior tubercle C5; b, infraglenoid tubercle; c, head of the first rib; D, protraction distance; E, elevation distance; x, upper triangle corner; y, lateral triangle corner; z, lower triangle corner; h, height; Cor, coracoid process).

Statistical analysis

Pearson's correlation coefficients (PCC) between the similarity measurements (DSC, JI, INI) and the Δ parameters were calculated. The result was one PCC for every similarity index with each morphometric parameter. PCC is a measure of the linear correlation between two variables, providing a value between +1 and −1, where 1 is the total positive correlation, 0 is no correlation, and −1 is the total negative correlation. The labelling systems were used to roughly categorize PCC, where correlation coefficients that are < 0.35 are generally considered to represent low or weak correlations, 0.35 to 0.67 modest or moderate correlations, and 0.68 to 1 strong or high correlations (Tayler et al. 1990).

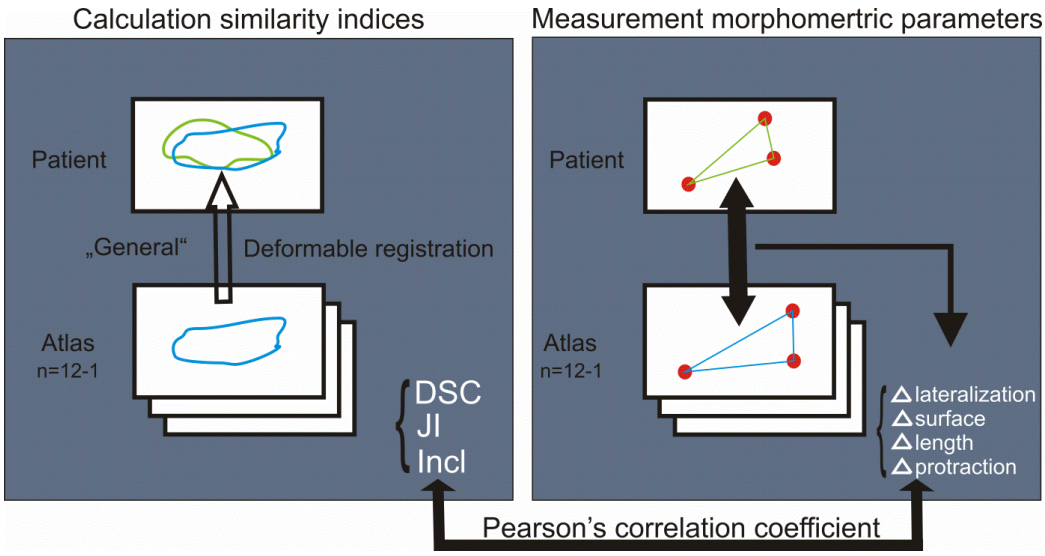


Fig. 14. Schematic illustration of the procedure for one patient. This procedure was repeated for every atlas as a patient (n=12). DSC, Dice similarity coefficient; JI, Jaccard Index; Incl, Inclusion index.

3. Results

Manuscript presentation

Manuscript 1

Brachial plexus 3D reconstruction from MRI with dissection validation – a baseline study for clinical applications

Joris Van de Velde, Stephanie Bogaert, Pieter Vandemaele, Wouter Huysse, Eric Achten, Joris Leijnse, Wilfried De Neve, Tom Van Hoof

Surg Radiol Anat 2015. Published

A1, IF: 1.33.

Abstract

Purpose: The present study aimed to establish a baseline for detailed 3D brachial plexus reconstruction from magnetic resonance imaging (MRI). Concretely, the goal was to determine the individual brachial plexus anatomy with maximum detail and accuracy achievable, as yet irrespective of whether the methods used could be economically and practically applied in the clinical setting.

Materials and methods: Six embalmed cadavers were randomly taken for MRI imaging of the brachial plexus. Detailed two-dimensional (2D) segmentation for all brachial plexus parts was done. The 2D brachial plexus segmentations were 3D reconstructed using Mimics® software. Then, these 3D reconstructions were anatomically validated by dissection of the cadavers. After finalising the cadaver experiments, brachial plexus MRI's were obtained in three healthy male volunteers and the same reconstruction procedure as in vitro was followed.

Results: A procedure was developed for brachial plexus 3D reconstruction based on MRI without the use of any contrast agent. Anatomical validation of six cadaver brachial plexus reconstructions showed high correspondence with the dissected brachial plexuses. Anatomical variations of the main branches were equally present in the 3D reconstructions generated. However, there were also some differences that related to the difference between the surface anatomy of the nerve and the internal nerve structure. In vivo, it was possible to reconstruct the complete brachial plexus in such a manner that normal appearing BP's were derived in a reproducible way.

Conclusions: This study showed that the described procedure results in accurate and reproducible brachial plexus 3D reconstructions.

Keywords: brachial plexus, three dimensional reconstruction, cadaver, anatomical validation, segmentation, procedure

Introduction

Detailed individual three-dimensional (3D) reconstructions of the brachial plexus (BP) based on MRI images have to our knowledge not been published. However, knowledge of individual 3D BP anatomy has direct clinical applications.

In radiotherapy of head-and-neck, breast and lung cancer, the BP is a structure at risk (SAR) [21]. Radiotherapists contour the target volumes (tumour) and the SARs on planning-computed tomography (CT). With poorly delineated BPs, incorrect SAR dose constraints may be set, with risk for radiation-induced brachial plexopathy [5, 11, 13].

Detailed 3D reconstructions may also improve preoperative planning for BP explorations, e.g., for resection of superior sulcus tumours [17].

In anaesthesiology, with BP nerve block injections, detailed 3D models may minimize the risk of pneumothorax or damage to the subclavian artery, subclavian vein or spinal cord [8, 9, 14, 18] and may lead to different needle insertion strategies.

More generally, accurate 3D BP reconstruction may enhance neurodynamic simulations in biomechanical modelling [10, 23].

For 3D BP reconstruction, magnetic resonance imaging (MRI) is the imaging modality of choice due to its superior soft tissue resolution and multiplanar capabilities. Continuous improvements in MRI scanners, coil and pulse sequence technology have resulted in the ability to perform routine, high-quality BP imaging. Different scanning protocols have been reported for BP visualization, but most publications [12, 24-26,] cite the T2 3D STIR (Short-Term Inversion Recovery) SPACE (Sampling Perfection with Application optimised Contrasts using different flip angle Evolution) sequence on a 3T MRI scanner as the best option for in vivo BP imaging. However, 3D BP reconstructions have rarely be obtained and when shown, only the distinction of the roots has been depicted. Individualized 3D reconstructions of the complete nerve root branching and merging patterns till the peripheral nerves have not yet been presented [8, 16].

Another possibly promising technique for BP visualisation is diffusion tensor imaging (DTI). However, motion and respiratory artefacts make BP DTI very challenging and so far only BP roots have been clearly visualised [12, 25].

In view of the limited past visualization results and with the direct long term aim of optimizing BP dose protection in radiotherapy, the present study aimed to establish a baseline for detailed 3D BP

reconstruction from MRI. Concretely, the goal was to determine the maximum detail and accuracy achievable, as yet irrespective of whether the methods used could be economically and practically applied in the clinical setting.

To this end, an MRI cadaver study was conducted, in contrast to the above cited studies which were performed in vivo. For establishing a baseline, cadaver MRI resolves important issues. First, the exclusion of movement artefacts, being in vivo always present from breathing and heartbeat. Second, the fact that the 3D reconstruction could be validated by dissection. Third, the possibility of experimenting with MRI scan parameters in time frames unacceptable in-vivo.

The aim of this study was to determine the individual BP anatomy on MRI with maximum detail and accuracy achievable and to set a baseline for detailed 3D BP reconstruction from MRI.

Materials and Methods

For the in vitro part of the study**,** MRI scans using a TIM 3T Magnetom Trio MRI system (Siemens®, Erlangen, Germany) were obtained from the right shoulders of six cadavers (age between 65 and 75 years; 4 women, 2 men) embalmed by the Thiel method [20]. Thiel embalming preserves tissue flexibility and full range of joint motion, which allowed positioning the bodies to the clinical standards. To obtain the scans, a 12-channel head coil, a 4-channel neck matrix coil, a 6-channel body matrix coil and a 24-channel spine matrix coil were integrated.

On the MRI table, cadavers were positioned supine, with neutral cervical spine and shoulder position and both arms alongside the body. A T1-weighted fat suppressed volume interpolated breath-hold examination (VIBE) sequence (voxel 0.8 × 0.8 × 0.8 mm), 320 mm field-of-view, 384 × 384 pixel, 0.8 mm slice increment, TE= 1.48 ms, TR= 4.12 ms and a scanning time of 32 min were used [23].

For the in vivo part, the resulting methods and MRI parameter settings were further used to obtain MRI scans of the right shoulders of 3 healthy male volunteers, age 25, 30 and 34 years. Cadaver scanning time of 32 minutes was reduced to 8 minutes for in vivo volunteers by decreasing the number of averages from six to two.

3D reconstruction

The obtained MRI datasets were imported into Mimics® 16.0 (Materialise, Leuven, Belgium), a post processing 3D software package. A full screen axial view was taken and on every slice the BP was manually segmented by an experienced anatomist with a one mm paint brush tool. Following each segmentation step, 3D reconstructions of the 2D segmentations were created to support the delineation process using the Mimics® 3D calculation function. The BP segmentation was performed up to where the BP terminal branches originate. Then a final BP 3D reconstruction was created. For enhanced visualisation purposes, smoothing of the 3D reconstruction was completed using the Mimics ‘smoothing’ function (iterations: 1, smoothing factor: 1) in combination with the ‘compensate shrinkage’ function.

2D segmentation procedures

BP segmentation from MRI is not obvious and we summarized our experiences into systematic procedures. While these segmentation procedures are part of the present methods, their description is extensive and is therefore included as supplementary material (online resource 1).

Dissection validation

After 3D reconstruction completion, the right cervical-shoulder region of each cadaver was dissected by two experienced anatomists according to an in-house protocol to preserve local topography [23]. The reconstruction validation was a qualitative comparison of all major branches of the plexus branching tree, which in each individual form a unique pattern, in the 3D BP reconstruction and the dissection findings [22]. In the course of these assessments it also became clear that branching nerves may already run largely separate for some distance within the nerve sheet before actually splitting. MRI, being cross-sectional, visualised the nerve splitting within the sheet, while dissection, relying on superficial observation of the nerve, established nerve splitting mainly by the division of the nerve sheet.

In vivo MRI scans

After finalising the cadaver experiments, BP MRI's were obtained in three healthy male volunteers using the same MRI parameter settings, coils and scan positions as used in vitro, but with following MRI sequence adaptations. The 32 minutes in vitro scanning time was in vivo reduced to 8 minutes by decreasing the number of averages from six to two. This resulted in a lower signal-to-noise ratio (SNR), but the components of the BP were still clearly visible. The aim of these scans was to determine how important in vivo motion artefacts were and to what degree in vivo MRIs would still allow detailed 3D BP reconstructions. Obviously, in vivo scans could not be validated by dissection, so there was no method to validate whether the reconstructions were realistic in detail. To obtain a general impression of the in vivo reconstruction precision, the inter-observer conformity was assessed. Therefore, 2 observers (experienced anatomists) independently segmented the BP of the same volunteer following the above described same segmentation procedure. The similarity of the branching pattern was compared. The time to complete one reconstruction was about 2 hours for both an in vivo dataset and a cadaver specimen.

Results

In vitro 3D BP reconstructions

In the qualitative comparisons of the BP roots, trunks, divisions, cords, and peripheral nerves in the 3D BP reconstructions and the dissected specimens, a high level of structural conformity was observed. Anatomical variations of the main branches were equally present in the 3D reconstructions generated; examples are given in Figure 1.

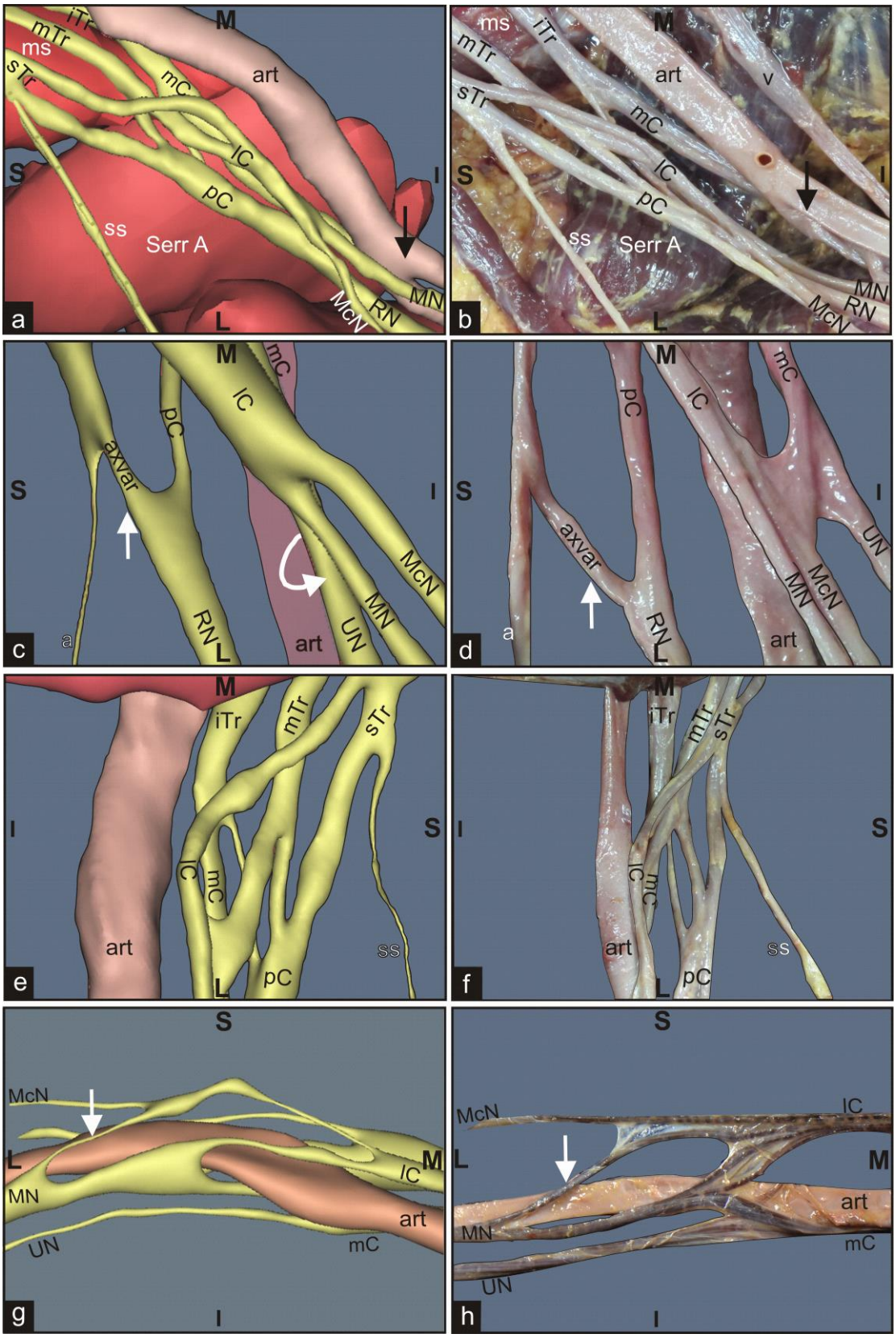


Fig. 1 3D-reconstructions of brachial plexuses are anatomically validated by cadaver dissection. Brachial plexus details of four different cadavers are presented. In the first cadaver **(a, b),** a bifurcation of the brachial artery is visible in both reconstruction **(a)** and dissection views **(b)** (black arrows). In the second cadaver **(c, d)**, an accessory axial nerve (axvar) connected with the posterior

cord is present in both reconstruction and dissection views (white arrows). The axillary artery (art) was not reconstructed, the median and musculocutaneous nerves are rotated around the ulnar nerve due to abduction-exorotation position of the shoulder in the dissection view (curved arrow). In the third cadaver **(e, f)** the branching pattern of the reconstruction and the dissection view are identical. In the fourth cadaver **(g, h)**, an accessory lateral contribution of the lateral cord to the median nerve is present in both reconstruction and dissection views (white arrows).

sTr, superior trunk; mTr, middle trunk; iTr, inferior trunk; pC, posterior cord; mC, medial cord; lC, lateral cord; ms, middle scalene; Serr A, serratus anterior; subsc, subscapular muscle; art, axillary artery; v, axillary vein; a, axillary nerve; RN, radial nerve; UN, ulnar nerve; MN, median nerve; McN, musculocutaneous nerve; ss, suprascapular nerve; M, medial; L, lateral; S, superior; I, inferior.

However, there were also some differences that clearly related to the difference between the surface anatomy of the nerve and the internal nerve structure. MRI, being volume (voxel) based, does not well distinguish thin layers, such as the nerve sheet. The MRI revealed that BP branches may consist for some length of two more or less distinct bundles, which may or may not remerge into more integrated single bundles. In the 3D reconstruction these would be segmented as a nerve splitting for some length and possibly remerging thereafter (Figure 2). In dissection, such a nerve would present as a single structure, as the internal nerve structure remained covered by the nerve sheath. Clearly, the thin nerve sheath lacked the cross-sectional area to be well visible in the MRI scans for reconstruction. Such nerve branch splitting artefacts in the 3D reconstructions were not rare (3/6 cadavers).

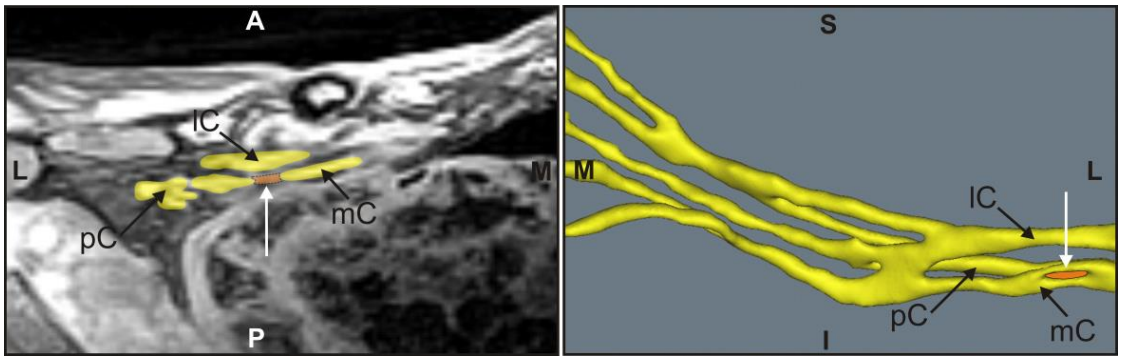


Fig. 2 Illustration of a nerve splitting artefact (white arrows). A splitting of the medial cord is visible (indicated in orange) which was not visible in the dissected brachial plexus. Left: 2D MRI illustration of the nerve splitting (axial view). Right: the same nerve splitting in the 3D reconstruction (posterior view). pC, posterior cord; mC, medial Cord; lC, lateral Cord; M, medial; L, lateral; S, superior; I, inferior; A, anterior; P, posterior.

Another difference between reconstructions and dissections is seen in terms of cross sectional area of the nerve branches. Nerves may be depicted larger or smaller due to the smoothing procedure

during 3D reconstruction. Despite the supplementary 'compensate shrinkage' function application, a slight alteration in nerve diameter is inevitable.

We also observed that differences in the location of the nerve bifurcation can occur. It is possible that in the 3D reconstruction, the bifurcation of a nerve branch can be depicted sooner than in the dissected cadaver (mean difference of 4,4 mm). This may be due to the ability to visualise the intra-neural branching pattern of nerves into different fascicular bundles on MRI as explained above [22]. An example can be seen in figure 3: the two different fascicles of the medial cord are reconstructed separately. In contrast, in the dissection only one medial cord is seen. Additionally, connective tissue is present between each nerve, as seen in Fig. 3 (white arrow). The density of all this connective tissue increases from the proximal to the distal portion of the BP [2]. As a consequence, there is a higher probability of imaging intra-neural bifurcations instead of macroscopic nerve bifurcations with MRI of the distal BP.

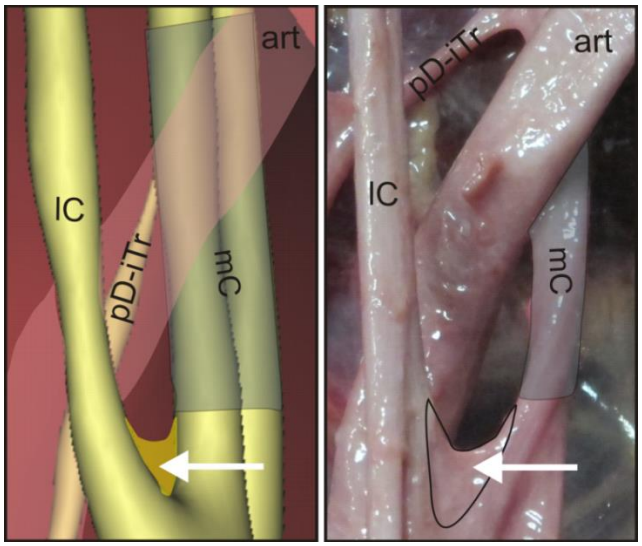


Fig. 3 Detail of brachial plexus reconstruction and corresponding cadaver dissection. The interneural connective tissue (white arrow) and the axillary artery were not reconstructed. (lC, lateral cord; mC, medial cord; pD-iTr, posterior division of the inferior trunk; art, axillary artery)

In vivo 3D BP reconstructions

Despite the lower signal-to-noise ratio due to a reduction in scanning time, a similar resolution was obtained in the in vivo scans. All components of the BP were still visible and could be 3D reconstructed (Figure 4). Only A limited decrease in resolution was seen in the BP division area. This was due to the movement artefacts of the nearby lung top. Nevertheless, it was possible to reconstruct the complete BP in such a way that normal appearing BPs were derived. Keeping in mind

that the in vivo reconstructions methodologically cannot be validated, there can only be assumed that the deviations from the actually present BPs should be small and the same would hold for the deviations from the total BP outline.

The 3D reconstructions of the 2 independent delineators following the same currently described procedure for reconstruction of the same in vivo MRI dataset showed an identical branching pattern.

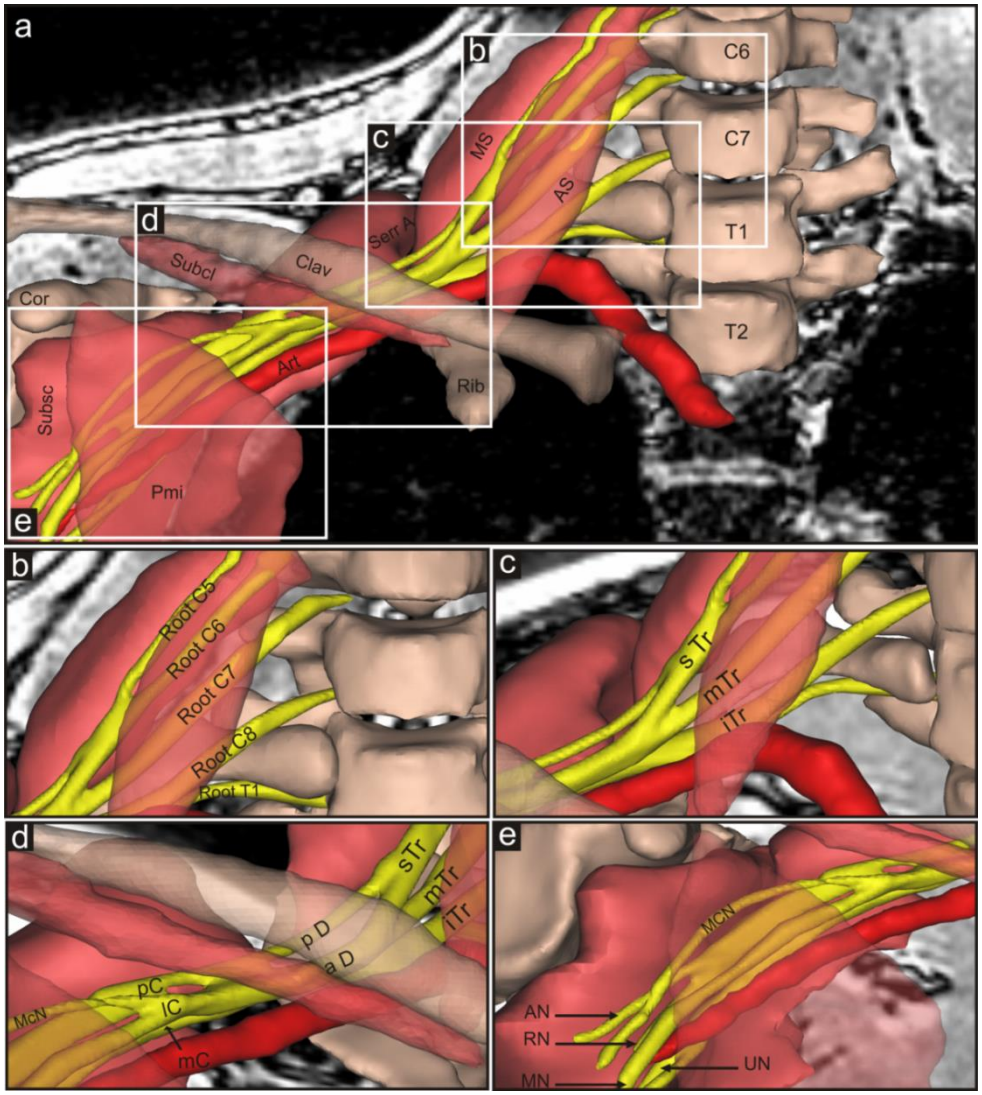


Fig. 4 MRI-based 3D reconstructions of the brachial plexus and surrounding structures of a healthy volunteer (AS, anterior scalene muscle; MS, middle scalene muscle; Serr A, serratus anterior muscle; Subcl, subclavian muscle; Pmi, minor pectoral muscle; Subsc, subscapular muscle; clav, clavicle; cor, coracoid process; Art, axillary artery) .

(a) shows an overview of the total course of the brachial plexus with the surrounding bones and muscles. **(b)** Shows an enlargement of the roots of the brachial plexus. (**c**) Shows the formation of the trunks (sTr, superior trunk; mTr, middle trunk; iTr, inferior trunk). **(d)** Shows the formation of the divisions and the cords (pD, posterior division of the superior trunk; aD, anterior division of the superior trunk; pC, posterior cord; lC, lateral cord; mC, medial cord) and **(e)** shows the formation of

the terminal branches (McN, musculocutaneous nerve; AN, axillary nerve; RN, radial nerve; MN, median nerve; UN, ulnar nerve).

Discussion

In this study, a procedure was developed for 3D BP reconstruction based on MRI without the use of any contrast agent. Anatomical validation of six cadaver BP reconstructions showed high correspondence with the dissected BPs. Anatomical variations of the main branches were equally present in the 3D reconstructions generated. However, there were also some differences that related to the difference between the surface anatomy of the nerve and the internal nerve structure.

In vivo, the same 3D reconstruction procedure was applied on three healthy volunteers. The in vivo application of the procedure showed a large inter-observer reliability and resulted in normal appearing BP 3D reconstructions.

The results indicated that in MRI based reconstructions the internal nerve structure dominates and may introduce nerve branch artefacts that would not be found in their macroscopic appearance during surgery. Clearly, the possibility of such reconstruction artefacts will hold for any structure covered by a thin sheath relative to the enclosed volumes.

To our knowledge, the observed artefacts are never described in literature.

The fact that no detailed BP 3D reconstructions are found in the literature may be due to the use of 2D MRI sequences in most publications [8, 15]. When 3D sequences are used, a detailed BP 3D reconstruction is never displayed. Only the BP roots are clearly 3 dimensionally depicted in those publications [8, 16]. This can be attributed to the complexity of recognising nerve branches in certain BP regions, such as the area of the divisions or the region of the cords, or perhaps the loss of information with slice thicknesses of more than 1 mm [6].

In the current study, a 3D T1-weighted fat suppressed VIBE sequence came out to be the best for BP visualisation through a ‘trial and error’ procedure. Several sequences and parameter settings were tested on cadavers for their capability to clearly visualise nerve tissue embedded in adipose tissue with high contrast. This spoiled gradient-echo sequence can acquire T1-weighted images with high quality and spatial resolution in a relatively short acquisition time. It was originally applied to

abdominal examinations and since then has gained wide acceptance for contrast-enhanced T1-weighted abdominal imaging [3,27]. In addition, fat-suppressed 3D VIBE has also been used in brain and chest imaging, whole-body tumour staging and in articular cartilage imaging [4]. To our knowledge, this MRI sequence was not previously used to visualize the BP.

The main clinical application for a detailed BP 3D reconstruction can be found in radiotherapy. In radiotherapy treatment planning, these 3D reconstructions can serve as template segmentations for automatic multi-atlas based BP segmentation of head-and-neck, lung top and breast cancer patients, for which a large amount of gold standard 3D BP segmentations is needed to derive the BP location on a de novo patient [22]. In radiology, detailed BP 3D reconstructions can help in understanding normal BP anatomy in situ, which can be important for highlighting BP pathologies, like e. g. nerve avulsions or thoracic outlet syndrome . In surgery, these 3D reconstructions can aid in preoperative planning for BP explorations in all patients were a thorough inside in the BP 3D anatomy is required. The anatomical details of the 3D reconstruction will provide surgeons with a huge amount of information to virtually plan BP explorations.

For practical reasons, the study was not conducted on fresh specimens, but on Thiel embalmed bodies [20]. Possibly, the embalming decreased inter-tissue water content differences, causing diminished MRI grey contrasts. However, this was not seen as a study limitation, but rather as a worst-case baseline approach: if scans could produce the detail presently obtained, scans of unembalmed or live tissues could be at least as good. Furthermore, the use of Thiel cadavers has previously provided very accurate models, resulting in high quality data for radiological, biomechanical and dosimetric studies [1, 7, 19, 23].

A practical limitation of the proposed 3D BP reconstruction technique is that it is based on an MRI sequence that is not in accordance with the sequences used in clinic. This means that an additional MRI sequence or an adapted sequence has to be used if BP 3D reconstruction is required. Also, the MRI sequence was tested on only 6 cadavers and 3 young, healthy volunteers, and not on patients suffering from plexopathies. In future, the feasibility of this technique and his cost effectiveness in patients with brachial plexopathies need to be investigated. At this moment, we assume that the proposed technique will be too time-consuming to be used for most clinical purposes. However, in future, a large database of individual BP 3D reconstructions can build up a statistical shape model of this structure. This statistical shape model can, to a great extent, contribute to an effective multi-atlas-based automatic BP segmentation method, which will ultimately signify a huge time saving for BP segmentation in clinical settings.

Conclusion

A detailed 3D BP reconstruction technique can be useful in several medical fields. Existing publications lack complete and detailed BP 3D reconstructions because no appropriate imaging sequences were used and detailed reconstructions have to be delineated manually which is difficult and time consuming, requiring both anatomical and computer software skills. A procedure for reconstruction of the complete BP was developed and anatomically validated by cadaver dissections in this study. Detailed 3D reconstruction of BP roots, trunks, cords and terminal branches was achieved with even depiction of small anatomical variations.

Conflict of interest

The authors have no conflicts of interest to report.

Acknowledgements

This research was supported by the Foundation against Cancer, grant 2012-200.

Ethical standards

The authors declare that the experiments comply with the current ethical laws of the state of Belgium.

References

1. Audenaert EA, Vigneron L, Van Hoof T, et al. (2011) In vitro validation and reliability study of electromagnetic skin sensors for evaluation of end range of motion positions of the hip. Med Biol Eng Comput 49:1405–1412

2. Bonnel F (1984) Microscopic anatomy of the adult human brachial plexus: an anatomical and histological basis for microsurgery. Microsurgery 5:107-117

3. Chandarana H, Block TK, Rosenkrantz AB, et al. (2011) Free-Breathing Radial 3D Fat-Suppressed T1-Weighted Gradient Echo Sequence A Viable Alternative for Contrast-Enhanced Liver Imaging in Patients Unable to Suspend Respiration. Invest Radiol 46:648-653

4. Chang G, Deniz CM, Honig S, et al. (2014) MRI of the Hip at 7T: Feasibility of Bone Microarchitecture, High-Resolution Cartilage, and Clinical Imaging. J Magn Reson 39:1384-1393

5. Chen AM, Hall WH, Li BQ, et al. (2011) Intensity-modulated radiotherapy increases dose to the brachial plexus compared with conventional radiotherapy for head and neck cancer. Brit J Radiol 84:58-63

6. Chhabra A, Thawait GK, Soldatos T, et al. (2013) High-Resolution 3T MR Neurography of the Brachial Plexus and Its Branches, with Emphasis on 3D Imaging. Am J Neuroradiol 34:486-497

7. De Crop A, Bacher K, Van Hoof T, et al. (2012) Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. Radiology 262:298-304

8. Kjelstrup T, Courivaud F, Klaastad O, et al. (2012) High-resolution MRI demonstrates detailed anatomy of the axillary brachial plexus. A pilot study. Acta Anaesthesiol Scand 56:914-919

9. Klaastad O, Smith HJ, Smedby O, et al. (2004) A novel infraclavicular brachial plexus block: The lateral and sagittal technique, developed by magnetic resonance imaging studies. Anesth Analg 98:252-256

10. Lien K, Morgan D, Delancey J, et al. (2005) Pudendal nerve stretch during vaginal birth: a 3D computer simulation. Am J Obstet Gynecol 192:1669-76

11. Lundstedt D, Gustafsson M, Steineck G, et al. (2012) Long-term symptoms after radiotherapy of supraclavicular lymph nodes in breast cancer patients. Radiother Oncol 103:155-160

12. Lutz AM, Gold G, Beaulieu C (2014) MR Imaging of the Brachial Plexus. Neuroimaging Clin N Am 24:91

13. Mutter RW, Lok BH, Dutta PR, et al. (2013) Constraining the brachial plexus does not compromise regional control in oropharyngeal carcinoma. Radiat Oncol 8(1):173

14. Raphael DT, McIntee D, Tsuruda JS, et al. (2005) Frontal slab composite magnetic resonance neurography of the brachial plexus - Implications for infraclavicular block approaches. Anesthesiology 103:1218-1224

15. Rehman I, Chokshi FH, Khosa F (2014) MR Imaging of the brachial plexus. Clin Neuroradiol 2:207–216

16. Rubin JA, Wesolowski JR (2011) Neck MR Imaging Anatomy. Magn Reson Imaging Clin N Am 19:457-473

17. Saji H, Kato Y, Shimada Y, et al. (2014) Three-dimensional multidetector computed tomography may aid preoperative planning of the transmanubrial osteomuscular-sparing approach to completely resect superior sulcus tumor. Gen Thorac Cardiovasc Surg Jan 25. [Epub ahead of print]

18. Sardesai AM, Patel R, Denny NM, et al. (2006) Interscalene brachial plexus block: Can the risk of entering the spinal canal be reduced? A study of needle angles in volunteers undergoing magnetic resonance imaging. Anesthesiology 105:9-13

19. Schuenke MD, Vleeming A, Van Hoof T, et al. (2012) A description of the lumbar interfascial triangle and its relation with the lateral raphe: anatomical constituents of load transfer through the lateral margin of the thoracolumbar fascia. J Anat 221:568–576

20. Thiel W (1992). Die Konservierung ganzer leichen in natürlichen farben. Ann Anat 174:185-195

21. Truong MT, Nadgir RN, Hirsch AE, et al. (2010) Brachial plexus contouring with CT and MR imaging in radiation therapy planning for head and neck cancer. Radiographics 30:1095-1103

22. Van de Velde J, Audenaert E, Speleers B, et al. (2013) An anatomically validated brachial plexus contouring method for Intensity Modulated Radiation Therapy planning. Int J Radiat Oncol 87:802-808

23. Van Hoof T, Gomes GT, Audenaert E, et al. (2008) 3D computerized model for measuring strain and displacement of the brachial plexus following placement of reverse shoulder prosthesis. Anat Rec 291:1173-1185

24. Vargas MI, Viallon M, Nguyen D, et al. (2010) Diffusion tensor imaging (DTI) and tractography of the brachial plexus: feasibility and initial experience in neoplastic conditions. Neuroradiology 52:237-245

25. Vargas MI, Viallon M, Nguyen D, et al. (2010) New approaches in imaging of the brachial plexus. Eur J Radiol 74:403-410

26. Viallon M, Vargas MI, Jlassi H, et al. (2008) High-resolution and functional magnetic resonance imaging of the brachial plexus using an isotropic 3D T2 STIR (Short Term Inversion Recovery) SPACE sequence and diffusion tensor imaging. Eur Radiol 18:1018-1023

27. Zheng Z-Z, Shan H, Li X (2010) Fat-Suppressed 3D T1-Weighted Gradient-Echo Imaging of the Cartilage With a Volumetric Interpolated Breath-Hold Examination. AJR Am J Roentgenol 194:W414-W419

Supplementary material: Procedures for brachial plexus 3D reconstructions.

For the segmentation of a standard brachial plexus (BP), its course was subdivided into four different anatomical regions. For each region, different surrounding reference structures were used for BP segmentation. The first region was the vertebral region (Fig. 1A, yellow). It started at the outcome of the BP roots in the intervertebral foramina, until their entrance in the scalene opening. In this region, the intervertebral foramina were used as a reference structure for the localisation of the BP. The second region was the interscalene region (Fig. 1A, blue). It started at the entrance in the scalene opening and ended at the entrance of the BP into the outlet between the serratus anterior and subclavian muscle. In the interscalene region, the anterior and middle scalene muscles were the reference structures. The third region was the subclavicular region (Fig. 1A, green). It started at the entrance of the BP into the serratus anterior-subclavian muscle outlet and ended with the course of the plexus behind the minor pectoral muscle. Here, the serratus anterior and subclavian muscles were the reference structures. The fourth region was the subcoracoid region (Fig. 1A, purple). In this region, the BP was located behind the minor pectoral muscle. The minor pectoral muscle was used as a reference structure, along with serratus anterior and subscapular muscles.

Procedures for BP segmentation

Each BP delineation starts in the axial plane. For determination of the correct vertebral levels, the sagittal plane may be more convenient.

In the vertebral region, five BP roots are leaving the intervertebral foramina. In the cervical spine, each root is situated above his eponymous vertebra. The root of C8 is situated between the pedicles of vertebra C7 and T1.

The segmentation starts with the delineation of root T1. To find this root, the slice where the head of the first rib is in contact with a vertebra is searched; this is the costovertebral articulation of vertebra T1. Root T1 is situated in the intervertebral foramen beneath vertebra of T1 (Fig. 1D). Root C8 is found in the intervertebral foramen above T1 (Fig. 1C). Then, to find the roots of C7, C6 and C5, scroll a few slices cranially. These roots appear as grey ellipsoid structures in front of the superior articular process of the underlying vertebra (Fig. 1B). They remain separate in their caudolateral course until they reach the space between the anterior and middle scalene muscles.

In the interscalene region, the three BP trunks are formed. When the root of C5 and C6 are followed in the scalene opening, the upper trunk is formed by fusion of these two roots in the lateral part of the scalene opening (Fig. 1C). The continuation of root C7 in the scalene opening is the middle trunk

(Fig. 1D). Root C8 and T1 fuse to form the inferior trunk (Fig. 1E). To find this fusion, the first slice were the top of the lung appears is searched. A few slices cranially, the inferior trunk is seen as a long and small ellipse in the axial plane just cranial to the top of the lung.

Region three (subclavicular region) is a complex and challenging region to segment. To gather more information in this region, experience learns it is better to continue with the segmentation of the more distally situated region four.

In region four (subcoracoid region), three BP cords are situated behind the minor pectoral muscle (Fig. 1H). More distally, these cords deliver the BP terminal branches in the space between the serratus anterior, subscapular and minor pectoral muscles (Fig. 1I, J, K). To find the three cords, it is advisable to start searching for the terminal branches. To find the musculocutaneous nerve, the coracobrachial muscle was located. The origin of this muscle is at the coracoid process. Starting from the origin, scroll down a few slices. After a variable amount of slices, a nerve structure is leaving the coracobrachial muscle; this is the musculocutaneous nerve (Fig. 1J). This nerve is to be followed by scrolling up until it reached one of the cords behind the minor pectoral muscle; this is the lateral cord. To find the radial nerve, more distally needs to be searched for the slice where a space appears between the coracobrachialis and subscapular muscles (Fig. 1K). The radial nerve is gradually running more and more posterior in this space. In contrast, the axillary nerve is seen here in only one or two slices as a small longitudinal structure running towards the quadrangular space (Fig. 1K). In case of MRI imaging of the most lateral BP regions (Fig. 2), supplementary reference structures can be used to locate the axillary nerve by scrolling down from the level of the humeral head. The supraspinatus and infraspinatus insertions appear anteromedial and posterolateral of the scapular spine respectively. If the image dataset allows complete lateral vision, the deltoid muscle enveloping the humeral column are indicated. By further scrolling down, the bulk of the long head of the triceps appears in the posterior deltoid region, reaching medially towards the inferior aspect of the glenoid cavity (infraglenoid tubercle) (Fig. 2). At that point, different features are seen over several slices in the caudal direction. At the anterior aspect of the humeral shaft, the medial crest of the tubercle, the bright conjoined tendon of the teres major and the latissimus dorsi muscle are depicted just posterior to the muscle fibres of the coracobrachialis and short head of the biceps. The area of the long head of the triceps drastically decreases and shifts posterolateral, forced by the emerging bulk of the teres major. At the same time, the muscle fibres of the teres major connect to their corresponding tendon and the medial aspect of the subscapular muscle abruptly moves posterolateral, leaving a small grey corridor in the posteromedial direction. From that moment, the grey area described as region 4 bordered by the anterior serratus, subscapular, coracobrachialis, major and minor pectoral muscles is united. In this grey corridor a longitudinal white line is depicted,

which is the axillary nerve crossing the quadrangular space and circumflexing the humeral shaft posterior to the subdeltoid region (Fig. 2). If the axillary nerve is followed in the anteromedial direction by scrolling back cranially up to the space behind the minor pectoral muscle, it merges into a grey ellipsoid area that can be labelled the posterior cord. From the moment the posterior cord and the lateral cord are identified, the only remaining cord in the space behind the minor pectoral muscle is the medial cord. This medial cord bifurcates more distally in a large medial branch, which is the ulnar nerve and a smaller medial contribution to the median nerve. This bifurcation is usually situated distally from the slice were the axillary artery is crossing the line between the medial and the lateral cords (Fig. 1J, K).

In region three (subclavicular region), the BP is passing between the subclavian and serratus anterior muscles. In this region, BP divisions are formed. For BP segmentation in this region, it is advisable to start in region two at the earlier situated three trunks for orientation. Each of the trunks bifurcate in an anterior and a posterior division. Scrolling from cranially to caudally, first the bifurcation of the superior trunk is seen just lateral to the scalene outlet. However, the formation of the anterior and posterior divisions of the superior trunk may occur in the interscalene region as well (Fig. 1D, E). More distally, the bifurcation of the middle trunk in an anterior and posterior division is seen (Fig. 1F). Finally, a few slices before the formation of the earlier situated three cords, the inferior trunk bifurcates (Fig. 1G). All three posterior divisions fuse to form the posterior cord. The anterior division of both the middle and superior trunk fuse to form the lateral cord (Fig. 1H, G), and the anterior division of the inferior trunk forms the medial cord (Fig. 1H). In this region all BP branches are closely packed; therefore, it is important that the more proximally and distally situated regions are segmented first, allowing easier identification of the remaining most difficult BP division parts. Starting from the BP cords, it is sufficient to segment all the small nerve branches that appear as little grey ellipses originating from these cords when scrolling up. When subsequently the 3D calculation function in Mimics processes the 3D visualisation of that region, the divisions are depicted veraciously, as seen in the anatomical validation procedure, and this complex anatomical region becomes apparent at that stage.

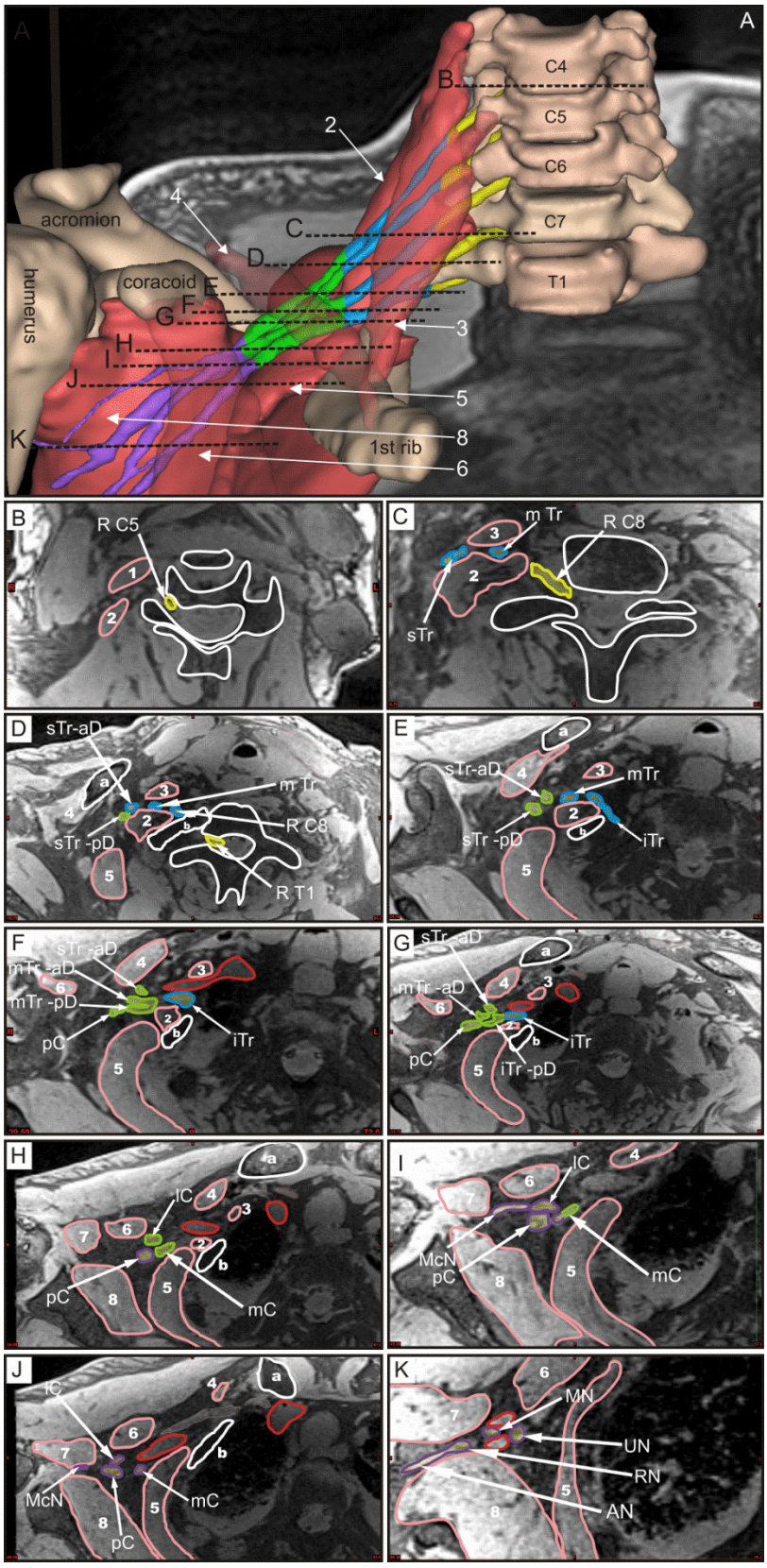


Fig. 1 (A) Shows the brachial plexus with surrounding muscles reconstructed in 3D.

Axial slices at 10 brachial plexus levels are shown in (B), (C), (D), (E), (F), (G), (H), (I), (J) and (K). Brachial plexus is shown in yellow in region 1, blue in region 2, green in region 3, purple in region 4. (R C5, root C5; R C7, root C7; R C8, root C8; R T1, root T1; sTr-aD, superior Trunk-anterior Division;

sTr-pD, superior Trunk-posterior Division; mTr-aD, middle Trunk-anterior Division; mTr-pD, middle Trunk-posterior Division; iTr-pD, inferior Trunk-posterior Division; pC, posterior Cord; lC, lateral Cord; mC, medial Cord; McN, Musculocutaneous Nerve; MN, Median Nerve; UN, Ulnar Nerve; RN, Radial Nerve; AN, axillary nerve). Surrounding muscles in pink (1, longus capitis muscle; 2, middle scalene muscle; 3, anterior scalene muscle; 4, subclavian muscle; 5, serratus anterior muscle; 6, minor pectoral muscle; 7, coracobrachial muscle; 8, subscapular muscle). Subclavian/axillary artery in red. Bones in white (a, clavicle; b, first rib)

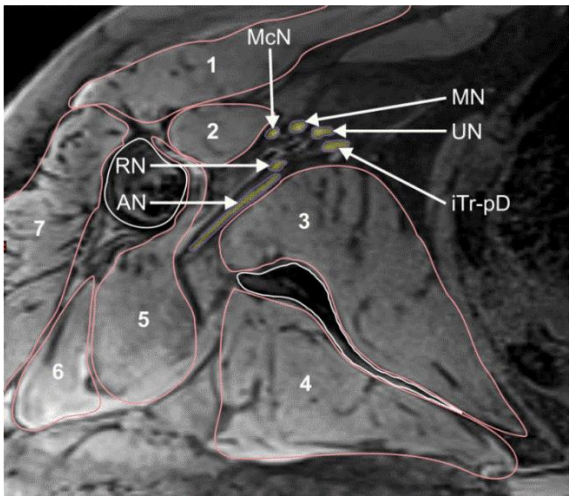


Fig. 2 MRI axial slice of the most lateral brachial plexus region, on the level of axillary nerve origin. The brachial plexus is depicted in purple (AN, axillary nerve; RN, radial nerve; McN, musculocutaneous nerve; MN, median nerve; UN, ulnar nerve; iTr-pD, inferior Trunk-posterior Division), muscles are pink (1, major pectoral muscle; 2, coracobrachial muscle; 3, subscapular musle; 4, infraspinatus muscle; 5, teres major muscle; 6, triceps muscle; 7, deltoid muscle)

Manuscript 2

Reliability and accuracy assessment of radiation therapy oncology group-endorsed guidelines for brachial plexus contouring

Joris Van de Velde, Tom Vercauteren, Werner De Gersem, Johan Wouters, Katrien Vandecasteele, Philippe Vuye, Frank Vanpachtenbeke, Katharina D'Herde, Ingrid Kerckaert, Wilfried De Neve, Tom Van Hoof

Strahlenther Onkol 2014;190:628-635. Published

A1, IF: 2.914

Abstract

Purpose: To validate the Radiation Therapy Oncology Group (RTOG)-endorsed guidelines for brachial plexus (BP) contouring by determining the intra- and interobserver agreement. Accuracy of the delineation process was determined using anatomically validated imaging datasets as a gold standard.

Materials and methods: Five observers delineated the right BP on three cadaver computed tomography (CT) datasets. To assess intraobserver variation, every observer repeated each delineation three times with a time interval of two weeks. The BP contours were divided into four regions for detailed analysis. Inter- and intraobserver variation was verified using the Computerized Environment for Radiation Research (CERR) software. Accuracy was measured using anatomically validated fused CT-magnetic resonance imaging (MRI) datasets by measuring the BP inclusion of the delineations.

Results: The overall kappa values were rather low (mean interobserver overall kappa: 0.29, mean intraobserver overall kappa: 0.45), indicating poor inter- and intraobserver reliability. In general, the kappa coefficient decreased gradually from the medial to lateral BP regions. The total agreement volume (TAV) was much smaller than the union volume (UV) for all delineations, resulting in a low Jaccard index (JI) (interobserver agreement, 0–0.124; intraobserver agreement, 0.004–0.636). The overall accuracy was poor, with an average total BP inclusion of 38%. Inclusions were insufficient for the most lateral regions (region 3, 21.5%; region 4, 12.6%).

Conclusion: The inter- and intraobserver reliability of the RTOG-endorsed BP contouring guidelines was poor. BP inclusion worsened from the medial to lateral regions. Accuracy assessment of the contours showed an average BP inclusion of 38%. For the first time, this was assessed using the original anatomically validated BP volume. The RTOG-endorsed BP guidelines have insufficient accuracy and reliability, especially for the lateral head-and-neck regions.

Key words: Brachial plexus, delineation, reliability, cadavers, validation

Introduction

Brachial plexopathy has been reported after high-dose radiation therapy of lung, breast and head-and-neck cancer (1–8). Accurate brachial plexus (BP) contouring is required to control the dose during the planning process. However, the existing guidelines for BP contouring endorsed by the Radiation Therapy Oncology Group (RTOG) are lacking validation (1,1,18,9). In this study we want to assess the validity of the RTOG-endorsed guidelines developed by Hall et al.(1) and modified by Yi et al.(18), by measuring intra- and inter-observer agreement and the accuracy of the delineation process using anatomically validated imaging datasets.

Clinical application of the RTOG-endorsed guidelines has revealed many inadequacies, especially in the use of anatomical reference points for delineation(1,1,18,5). Anatomical data indicate that these reference points are too variable in their topographical relationship with the BP(0). Furthermore, the guidelines are elaborated only for a limited number of slice levels(1,1,18,9). Therefore, when non-indicated slice levels need to be delineated, key information is lacking. Also, important details are not clearly defined and are thus subject to interpretation.

Yi et al.(18) have stated that the RTOG-endorsed BP contouring atlas provides a reliable set of guidelines for consistent BP contouring and a robust tool for accurate BP delineation. However, these findings were derived from a procedure based on mean BP volumes, ranges of maximal doses and ranges of BP volumes receiving 60 Gy or greater for five patients across three observers(18). These parameters are not indicative of reliability. Moreover, they did not perform appropriate statistical analysis to determine the reproducibility of the guidelines.

Currently, accuracy assessment is obtained by comparison of the delineations with a ‘gold’ standard, which to date is only derived from a mathematical average contour, consensus reading or expert opinion(11–15). Yi et al.(18) performed accuracy assessment subjectively, with visual verification of the contours without the use of a gold standard. A more detailed explanation of their methodology is not found, which makes it impossible to draw conclusions about delineation accuracy.

The purpose of the present study is to assess inter- and intra-observer reliability of the RTOG-endorsed BP contouring method(5). The accuracy of the delineations will be determined using anatomically validated cadaver imaging datasets as gold standard(0). Cadavers provide the opportunity for long MRI-scanning times, unlimited exposure to radiation, optimal standardization of body posture between CT and MRI and exclusion of movement, flow and breathing artifacts.

Materials and methods

For the development of anatomically validated imaging datasets, three randomly selected cadavers (three females; average age, 72 years) were embalmed according to the Thiel method(16), which is characterized by flexible soft tissues and joints, full range of motion and maintenance of tissue integrity(16,8). The use of Thiel cadavers has previously provided accurate models, resulting in high-quality data for biomechanical and dosimetrical studies(8–10).

Cadavers were positioned on a carbon plate in standard head-and-neck IMRT position (supine, neutral cervical spine and shoulder position, with both arms positioned alongside the body). The head, cervical spine and upper quadrant were immobilized using a thermoplastic material (Orfit®). First, an MRI-dataset was obtained using a 3-T high-field TIM TRIO MRI scanner (Siemens®) with a T1-weighted fat-suppressed VIBE sequence (0.83 × 0.83 × 0.83 mm voxel, 320 mm field-of-view, 384 × 384 pixel, 0.8-mm slice thickness and 32-min scan time)(0,12). The cadaver was then transported on the carbon plate to the IMRT-planning facility. There, the right side was imaged with a helical CT-scan (Toshiba®, Aquilion) using the following scanning parameters: 120 kVp, 300 mA, 1-mm slice increment, 502.78 mm field-of-view and 512 × 512 pixels.

CT and MRI datasets were separately imported into the 3D software package Mimics 15.0®, resulting in separate CT- and MRI-project files. The BP was manually segmented and 3D reconstructed in the MRI-project file (0,12). Bony structures were reconstructed from the CT-project files. The 3D BP from the MRI-project file was imported into the corresponding CT-project files, and was appropriately positioned by applying the transformation matrix to generate the ‘enhanced CT-project’ file.

The right cervical-shoulder region of each cadaver was then dissected according to an in-house protocol which maximally preserved local topography(12). The 3D-reconstructions of the BP were compared with the corresponding regions in the dissected cadaver(0). This resulted in the anatomical validation of the enhanced CT-project file used for assessment of the delineations’ accuracy.

Five observers (two experienced radiation oncologists, three experienced dosimetrists) participated in the BP delineation process. In Mimics®, every observer contoured the right BP on the non-enhanced CT for the three cadavers on three different days in random order, using the RTOG-endorsed BP contouring guidelines(18) (Table 1). This procedure was repeated three times per cadaver for each observer with a 2-week time interval.

Table 1. The RTOG-endorsed brachial plexus contouring guidelines

1. Identify and contour C5, T1, and T2.
2. Identify and contour the subclavian and axillary neurovascular bundle.
3. Identify and contour anterior and middle scalene muscles from C5 to insertion onto the first rib.
4. To contour the brachial plexus OAR use a 5-mm diameter paint tool.
5. Start at the neural foramina from C5 to T1; this should extend from the lateral aspect of the spinal canal to the small space between the anterior and middle scalene muscles.
6. For CT slices, where no neural foramen is present, contour only the space between the anterior and middle scalene muscles.
7. Continue to contour the space between the anterior and middle scalene muscles; eventually the middle scalene will end in the region of the subclavian neurovascular bundle.
8. Contour the brachial plexus as the posterior aspect of the neurovascular bundle inferiorly and laterally to one to two CT slices below the sternoclavicular joint and one to two ct slices superior to the level of the top of the aortic arch and the inferior aspect of the glenohomeral joint. The contour should have an approximate transverse length of 3 to 4 cm at this level.
9. The first and second ribs serve as the medial and approximate lateral limit of the OAR contour. The contour should not pass inferiorly to the level of the second rib.

The BP delineations were then divided into four regions for detailed analysis. Region 1 extended from the exit of the BP through the intervertebral foramina until its entrance in the scalene opening. Region 2 started at the entrance of the scalene opening and ended at the entrance between the subclavius and serratus anterior muscles. In region 3, the BP was defined between the subclavius and serratus anterior muscles. In region 4, the BP was bordered between the minor pectoral, subscapular and serratus anterior muscles.

Statistical analysis

Computerized Environment for Radiation Research (CERR) was used to analyze the reliability of the delineations. CERR is an open-source Matlab-based radiation therapy planning analysis tool. Both the inter- and intraobserver reliability were characterized using the statistical parameters described below.

In each region, contours were compared for agreement by calculating the apparent volume overlap, which is the average agreement probability by which a voxel is selected by the observers

(interobserver agreement) or repeatedly by one observer (intraobserver agreement). This value was corrected for agreement by chance by using the generalized kappa statistics(21). Kappa statistics assume values from +1 (perfect agreement) to 0 (no agreement above chance) or -1 (complete disagreement). According to the Landis and Koch criteria, kappa values are interpreted as follows: 0, poor agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement(43–24).

We also calculated the Jaccard similarity index (JI), as the ratio of the total agreement volume (TAV; the intersection volume of the delineations) and the entire union volume of the delineations (UV). The JI is situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

To measure the interobserver reliability, each observer's first delineation was imported into CERR. This was repeated for each cadaver. Afterwards the observer's second and third delineations were imported in a similar manner. The parameters were expressed per region for each cadaver, as the mean value of the three delineations for each observer (Table 2).

Table 2. Summary of inter-observer statistics

Cadaver number	Region	Overall kappa	TAV (cm^3)	UV (cm^3)	JI (TAV/UV)
1	1	0.504	1.848	15.915	0.116
	2	0.386	0.181	25.621	0.070
	3	0.175	0	5.835	0
	4	0.025	0	11.840	0
2	1	0.442	1.768	23.574	0.070
	2	0.357	0.367	20.744	0.030
	3	0.252	0.002	12.346	0.001
	4	0.069	0	12.745	0
3	1	0.494	1.917	17.098	0.112
	2	0.541	2.477	19.969	0.124
	3	0.275	0.009	3.9415	0.002
	4	0.054	0	9.877	0
mean		**0.298**	**0.714**	**14.959**	**0.038**

Abbreviations: TAV = Total Agreement Volume; UV= Union Volume; JI= Jaccard index

To measure the intraobserver reliability, the three delineations of each observer for a single cadaver were imported into CERR. This was repeated for each cadaver. To calculate intraobserver reliability, the different parameters were expressed per region as the mean for the three cadavers for each observer (Table 3).

Table 3. Summary of intra-observer statistics

Observer	Region	Overall kappa	TAV (cm^3)	UV (cm^3)	JI (TAV/UV)
1	1	0.694	6.531	14.597	0.447
	2	0.531	3.028	14.915	0.203
	3	0.298	0.552	8.210	0.067
	4	0.136	0.031	6.267	0.004
2	1	0.599	2.960	8.545	0.346
	2	0.659	4.247	10.919	0.388
	3	0.392	0.400	5.716	0.069
	4	0.058	0.167	4.396	0.037
3	1	0.639	7.091	16.031	0.442
	2	0.370	1.605	7.010	0.228
	3	0.663	1.350	2.736	0.493
	4	0.200	0.073	4.335	0.016
4	1	0.623	5.521	14.382	0.384
	2	0.457	3.303	14.866	0.222
	3	0.495	2.315	7.151	0.323
	4	0.140	0.131	8.520	0.015
5	1	0.773	7.684	12.075	0.636
	2	0.756	8.323	16.028	0.519
	3	0.176	0.183	1.993	0.091
	4	0.427	1.331	11.164	0.119
Mean		**0.455**	**2.842**	**9.493**	**0.253**

Abbreviations: TAV = Total Agreement Volume; UV= Union Volume, JI= Jaccard index

For accuracy assessment of the delineations, each observer's delineation was imported into the corresponding enhanced CT-project file and compared with the anatomically validated BP

contour as the gold standard (Fig. 1). Therefore inclusion, defined as the ratio of the intersection volume of the actual BP and the delineation to the actual BP volume, was measured for each BP delineation.

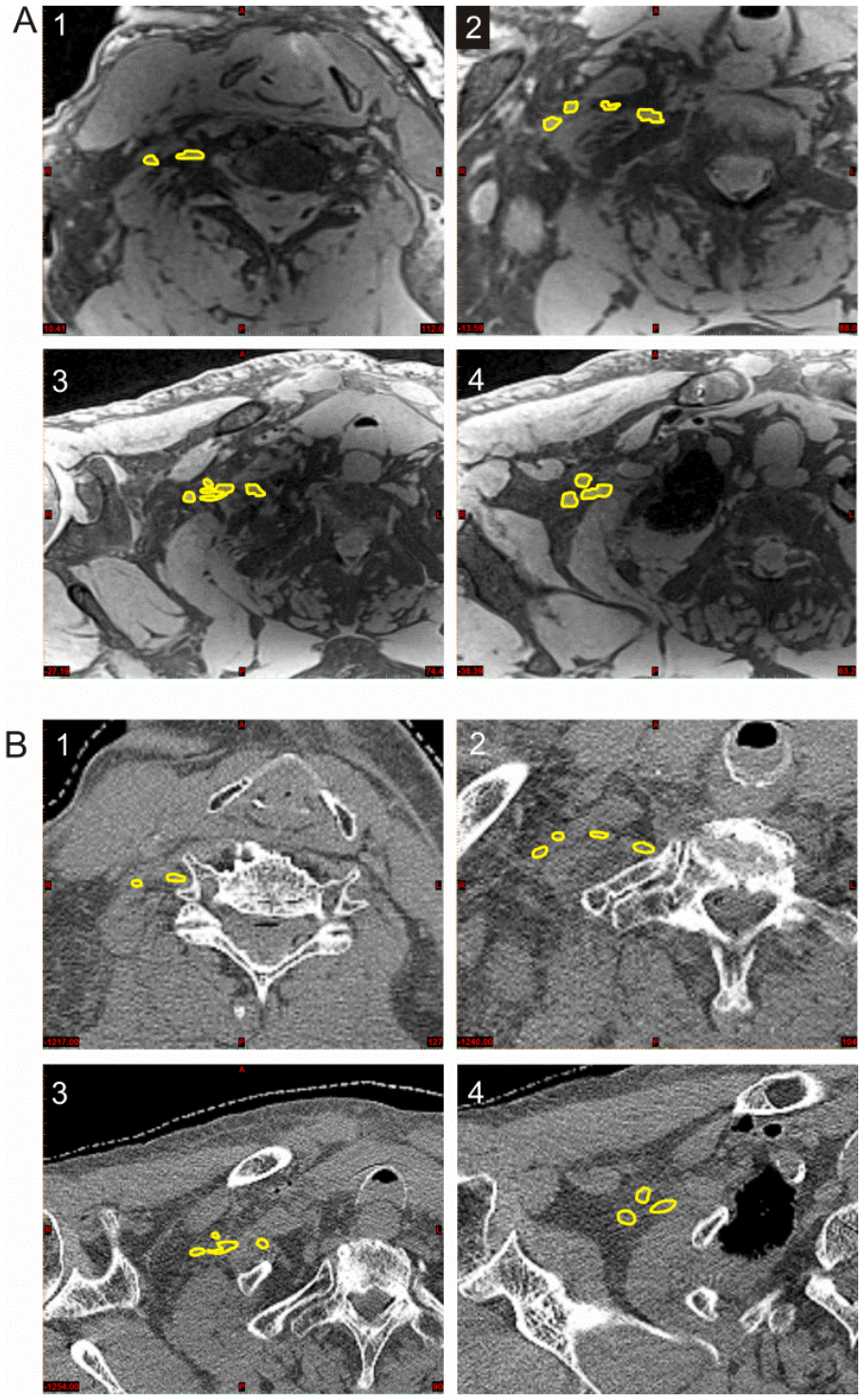


Fig. 1. The anatomically validated BP contours on MRI (A) and CT (B) in four different BP regions (1-4).

Results

The interobserver reliability statistics are summarized in Table 2 per cadaver per region as an average of the three sets of delineations performed by each observer. We observed perfect agreement in the sequence of the reliability parameters (overall kappa and JI; Table 2). BP delineations showed moderate interobserver agreement in regions 1 and 2 and fair-to-slight interobserver agreement in regions 3 and 4 (Fig. 2A). The TAV was much smaller than the UV for all the delineations, resulting in a low JI (mean, 0.038).

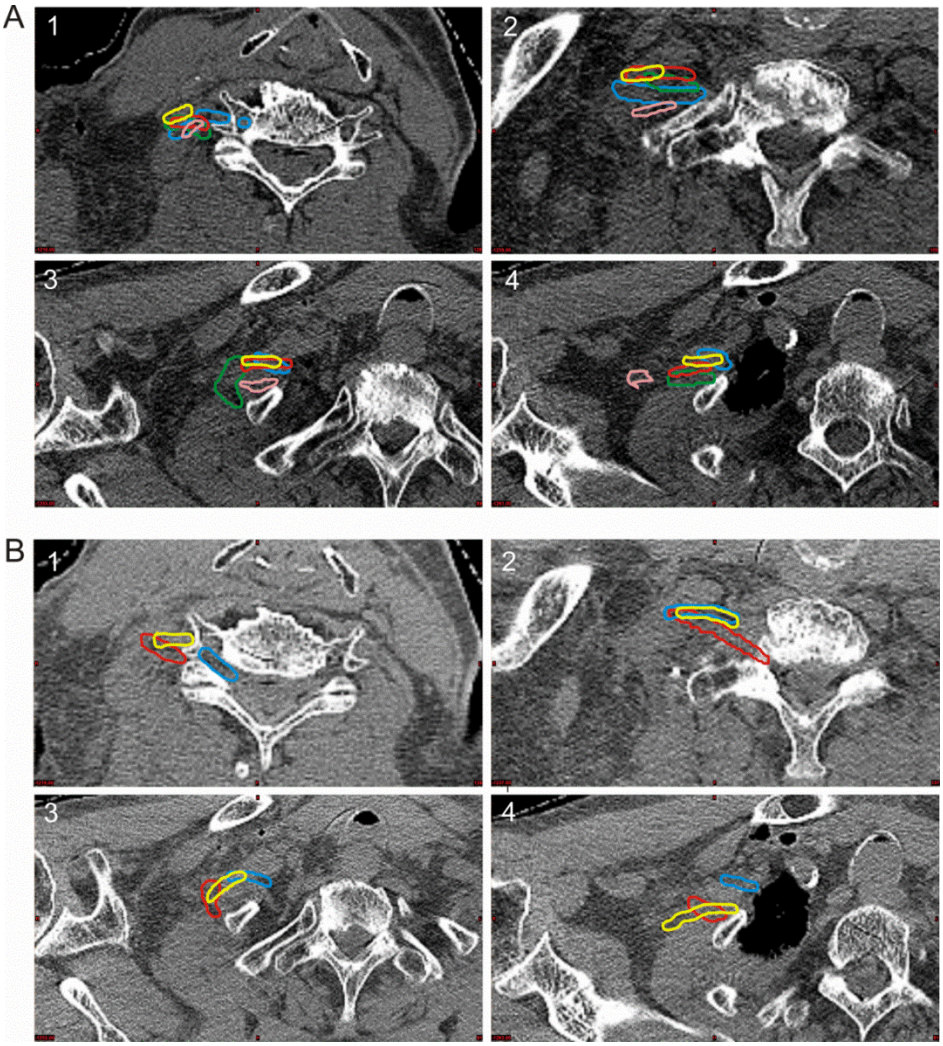


Fig. 2. (A) The delineations of five different observers on one cadaver CT-dataset (yellow, pink, red, green and blue) in one axial slice in each BP regions (1–4).

(B) Three delineations of one observer on a cadaver CT-dataset (red, yellow and blue) in one axial slice in each BP regions (1–4).

The intraobserver reliability statistics for each observer per region were processed as the average of the three cadavers (Table 3). Agreement was found in the sequence of the reliability

parameters (overall kappa and JI). BP delineations following the RTOG-guidelines showed only moderate-to-substantial intraobserver agreement in regions 1 and 2 and slight-to-fair intraobserver agreement in regions 3 and 4 (Fig. 2B). The JI was small for all delineations, and was slightly higher in regions 1 and 2 than in regions 3 and 4.

Among the five observers, the mean BP inclusion in regions 1–4 were 66.9%, 51%, 21.5% and 12.6%, respectively (Fig. 3). The total BP inclusion (in all four regions) for one observer was merely 38% (Fig. 3). The BP inclusion percentages showed a clear downward trend from the medial to lateral regions (Fig. 4).

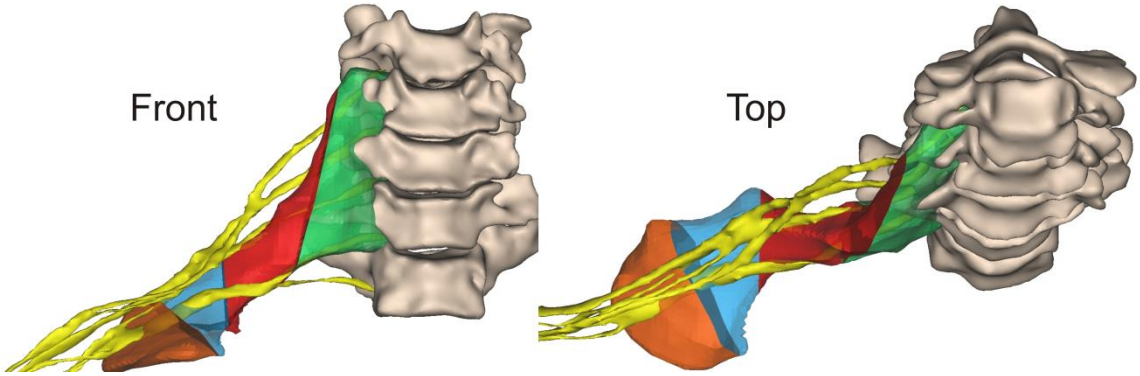


Fig. 3. Brachial plexus delineation by one observer on a cadaver dataset. In this example, the total delineation (region 1, green; region 2, red; region 3, blue; region 4, orange) shows 38.43% BP (yellow) inclusion.

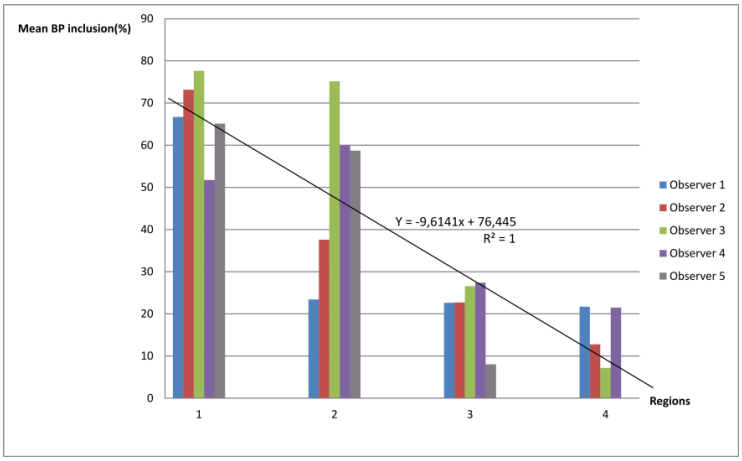


Fig. 4. Mean BP inclusion (%) per observer over the three CT-datasets illustrated per region (1–4). A downward trend is seen: line from region 1 to region 4 (mean slope: -9.61, range: -6.7 to -13; mean R^2: 0.81, range: 0.630 to 0.916).

Discussion

The purpose of this study was to determine the inter- and intraobserver agreement and accuracy of the RTOG-endorsed guidelines for BP contouring. For accuracy assessment, we used anatomically validated imaging datasets as a gold standard.

We found that the overall kappa values were rather low (mean interobserver overall kappa: 0.29, mean intraobserver overall kappa: 0.45), indicating insufficient reliability. In general, the kappa coefficient decreased gradually from region 1 to region 4.

The mean JI for both the inter- and the intra-reliability study was also low (mean interobserver JI: 0.038, range: 0.0–0.124; mean intraobserver JI: 0.253, range: 0.004–0.636). There was strong agreement among the reliability parameters (kappa and JI).

The overall accuracy was poor, with an average inclusion of 38% for the entire BP and with even lower percentages for the subclavicular and subcoracoid regions (21.5% in region 3 and 12.6% in region 4). The accuracy of the delineations decreased from the medially to laterally situated BP regions (Fig. 4).

In the first region, the average BP inclusion was 67%. A substantial part of the BP (33%) was not covered because none of the observers delineated the area around the T1 root. The guidelines provide no information about the delineation of this root. Instead, major parts of the scalene muscles are included in the entire first region. The mean overall kappa for interobserver reliability was ‘moderate’ (0.48), and the mean intraobserver overall kappa was ‘substantial’ (0.67). This region represented the highest accuracy and inter- and intraobserver agreement of all regions. This may be attributable to the close anatomical relationship in this region between the BP roots and the intervertebral foramina, which are used as reference structures in the guidelines. This relationship is strong, with only a small degree of anatomical variation.

The second region showed a mean BP inclusion of approximately 50% and a ‘moderate’ overall kappa for inter- and intraobserver reliability (0.43 and 0.55, respectively). The BP parts not included by the guidelines were the upper and middle trunks, extending laterally from between the anterior and middle scalene muscles, in proximity to their costal insertions. Since the guidelines only instruct to contour the space between the scalene muscles, a major part of the BP is ignored, leaving the upper and middle trunks completely unprotected. Instead, the delineations will include large areas of the scalene muscles. The moderate inter- and intraobserver agreement can be attributed to the incorrect identification of the scalene muscles in this region.

In regions 3 and 4, the BP inclusion was only 23% and 13%, respectively. The delineations covered large parts of the surrounding structures, i.e., the scalene muscles, serratus anterior muscle and non-neural elements of the subclavicular and subcoracoid region. Both inter- and intraobserver agreement were 'fair' in region 3 (0.23 and 0.38, respectively) and 'slight' in region 4 (0.05 and 0.19, respectively). These low inclusion percentages and agreements in general are the consequence of various issues.

First, the guidelines refer to the subclavian neurovascular bundle. The borders of this anatomical structure are difficult to distinguish on CT, and the guidelines only prescribe to contour the posterior aspect of this bundle. In fact, in these regions, the BP is actually located more to the lateral side of the bundle.

Second, the reference structures provided in the guidelines to indicate the inferior level of the BP contour in these regions (sternoclavicular joint, aortic arch, glenohumeral joint and second rib) are too variable in relation to the BP topography to use as reference structures. And finally, the ambiguous formulation of the guidelines and incorrect use of terms for anatomical orientation make the guidelines difficult to interpret correctly. All this will inevitably lead to erroneous delineations leaving the fascicles partly and often completely unprotected.

The current findings are in contradiction with the reports of Hall et al.(1) and Yi et al.(18). Both studies state that the BP contouring atlas developed by Hall et al. provides a reliable set of guidelines for consistent BP contouring and a robust tool for accurate BP delineation(1,18). However, their conclusions are not substantiated by any statistical analysis. Neither accuracy nor inter- and intraobserver reliability were studied in these reports(1,18). Only visual assessment of the contours was performed without any further processing. In the current study, we used anatomically validated imaging datasets as a gold standard for accuracy assessment. In the BP delineation studies of Truong et al.(1) and Kong et al.(5), no tests for accuracy or reliability were included.

In the clinic, low accuracy of anatomical contouring will contribute to uncertainties in radiation treatment planning, which would be even larger than those caused by set-up errors and organ motion for some tumor sites(26). High inter- and intraobserver variability in turn will lead to even less accurate BP delineation, further increasing the risk of brachial plexopathy(1,18). As high-dose regions are often present within millimeters of the delineated BP(18), inaccurate contouring may easily result in unwanted high doses to the BP.

The concept of anatomical validation has never been used in the evaluation process of delineation guidelines. In the past, the choice of a gold standard contour was a mathematical average contour, consensus reading or expert opinion(10–15). In the literature, a gold standard that outlines the true extent of the object is lacking, which makes it impossible to draw conclusions about the absolute accuracy of contours(26,27).

The method proposed in this study introduces a gold standard procedure that is not interpretation-based, but grounded on segmentation verification by cadaver dissection. This method has never been used in radiotherapy and could mean the optimization of other gold standard-development procedures, especially in cases of poor CT-visibility of structures, when expert opinion, consensus reading or mathematical average contour calculation is inadequate to develop an optimal gold standard contour.

Conclusion

The inter- and intraobserver reliability of the RTOG-endorsed BP contouring guidelines are poor, with decreasing quality of delineation from medial to lateral BP regions. The RTOG-endorsed BP guidelines have insufficient accuracy and reliability, resulting in erroneous delineations that in turn lead to suboptimal planning and protection of the BP as an organ at risk. Undesirable referral of BP parts to vertebral or other median structures and the use of unclear terminology are the main causes of the guidelines' poor quality.

In this study, for the first time IMRT contouring was assessed using an original anatomically validated BP volume. We recommend more intensive study to develop BP contouring guidelines based on computer-aided techniques and gold standard optimization.

On behalf of all authors, the corresponding author states that there is no conflict of interest.

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975 (in its most recently amended version).

References

1. Hall WH, Guiou M, Lee NY (2008) et al. Development and validation of a standardized method for contouring the brachial plexus: Preliminary dosimetric analysis among patients treated with IMRT for head-and-neck cancer. Int J Radiat Oncol Biol Phys 72:1362-1367.
2. Truong MT, Nadgir RN, Hirsch AE (2010) et al. Brachial plexus contouring with CT and MR imaging in radiation therapy planning for head-and-neck cancer. Radiographics 30:1095-1103.
3. Amini A, Jang J, Williamson R (2012) et al. Dose constraints to prevent radiation-induced brachial plexopathy in patients treated for lung cancer. Int J Radiat Oncol Biol Phys 82:391-398.
4. Lundstedt D, Gustafsson M, Steineck G (2012) et al. Long-term symptoms after radiotherapy of supraclavicular lymph nodes in breast cancer patients. Radiother Oncol 103:155-160.
5. Yi SK, Hall WH, Mathai M (2012) et al. Validating the rtog-endorsed brachial plexus contouring atlas: An evaluation of reproducibility among patients treated by intensity-modulated radiotherapy for head-and-neck cancer. Int J Radiat Oncol Biol Phys 82:1060-1064.
6. Atean I, Pointreau Y, Barillot I (2012) et al. Volumes de délinéation dans le traitement des cancers du sein : volumes cibleset organes à risque. Cancer/ Radiothérapie 16:485-492.
7. Chen AM, Hall WH, Li J (2012) et al. Brachial plexus-associated neuropathy after high-dose radiation therapy for head-and-neck cancer. Int J Radiat Oncol Biol Phys 84(1):165-9.
8. Platteaux N, Dirix P, Hermans R (2010) et al. Brachial Plexopathy after Chemoradiotherapy for Head-and-neck Squamous Cell Carcinoma. Strahlenther Onkol 186:517–20
9. Kong FM, Ritter T, Quint DJ (2011) et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. Int J Radiat Oncol Biol Phys 81:1442-1457.
10. Van de Velde J, Audenaert EA, Speleers B (2013) et al. An anatomically validated brachial plexus contouring method for intensity modulated radiation therapy planning. Int J Radiat Oncol Biol Phys 87:802-808
11. Kepka L, Bujko K, Garmol D (2007) et al. Delineation variation of lymph node stations for treatment planning in lung cancer radiotherapy. Radiother Oncol 85:450-455.
12. Petersen RP (2007) et al. Target volume delineation for partial breast radiotherapy planning: Clinical characteristics associated with low interobserver concordance. Int J Radiat Oncol Biol Phys 69:41-48.
13. Song WY, Chui B, Bauman GS (2006) et al. Prostate contouring uncertainty in megavoltage computed tomography images acquired with a helical tomotherapy unit during image-guided radiation therapy. Int J Radiat Oncol Biol Phys 65:595-607.
14. Breen SL, Publicover J, De Silva S (2007) et al. Intraobserver and interobserver variability in gtv delineation on fdg-pet-ct images of head-and-neck cancers. Int J Radiat Oncol Biol Phys 68:763-770.

15. Simmat I, Georg P, Georg D (2012) et al. Assessment of accuracy and efficiency of atlas-based autosegmentation for prostate radiotherapy in a variety of clinical conditions. Strahlenther Onkol 188:807–815
16. Thiel W (1992). Die Konservierung ganzer leichen in natürlichen farben. Ann Anat 174: 185-195.
17. De Crop A, Bacher K, Van Hoof T (2012) et al. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. Radiology 262:298-304.
18. Audenaert EA, Vigneron L, Van Hoof T (2011) et al. In vitro validation and reliability study of electromagnetic skin sensors for evaluation of end range of motion positions of the hip. Med Biol Eng Comput 49:1405-1412.
19. Schuenke MD, Vleeming A, Van Hoof T (2012) et al. A description of the lumbar interfascial triangle and its relation with the lateral raphe: anatomical constituents of load transfer through the lateral margin of the thoracolumbar fascia. J Anat 221:568–576.
20. Van Hoof T, Gomes GT, Audenaert E (2008) et al. 3D computerized model for measuring strain and displacement of the brachial plexus following placement of reverse shoulder prosthesis. Anat Rec 291:1173-1185.
21. Fleiss JL, Levin BA, Paik MC (2003). Statistical methods for rates andproportions. 3rd ed. Hoboken, NJ: Wiley.
22. Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. Biometrics 33:159–174.
23. Lawton CAF, Michalski J, El-Naqa I (2009) et al. Variation in the definition of clinical target volumes for pelvic nodal conformal radiation therapy for prostate cancer. Int J Radiat Oncol Biol Phys 74:377-382.
24. Ost P, De Meerleer G, Vercauteren T (2011) et al. Delineation of the postprostatectomy prostate bed using computed tomography: interobserver variability following the EORTC delineation guidelines. Int J Radiat Oncol Biol Phys 81:143-149.
25. Hoyte L,Ye W, Brubaker L (2011) et al. Segmentations of MRI Images of the Female Pelvic Floor: A Study of Inter- and Intra-reader Reliability. JMRI ; 33: 684-691.
26. Jameson MG, Holloway LC, Vial PJ (2010) et al. A review of methods of analysis in contouring studies for radiation oncology. JMIRS 54: 401-410.
27. Warfield SK, Zou KH, Wells WM (2008). Validation of image segmentation by estimating rater bias and variance. Phil. Trans. R. Soc. A. 366: 2361–2375.

Manuscript 3

An anatomically validated brachial plexus contouring method for intensity modulated radiation therapy planning

Joris Van de Velde, Emmanuel Audenaert, Bruno Speleers, Tom Vercauteren, Thomas Mulliez, Pieter Vandemaele, Eric Achten, Ingrid Kerckaert, Katharina D'Herde, Wilfried De Neve, Tom Van Hoof.

Int J Radiat Oncol Biol Phys 2013;87:802-808. Published

A1, IF: 4.176.

Abstract

Purpose: To develop contouring guidelines for the brachial plexus (BP) using anatomically validated cadaver datasets. Magnetic resonance imaging (MRI) and computed tomography (CT) were used to obtain detailed visualizations of the BP region, with the goal of achieving maximal inclusion of the actual BP in a small contoured volume, while also accommodating for anatomical variation.

Methods and materials: CT and MRI were obtained for eight cadavers positioned for intensity-modulated radiation therapy (IMRT). Three-dimensional reconstructions of soft tissue (from MRI) and bone (from CT) were combined to create eight separate enhanced CT-project files. Dissection of the corresponding cadavers anatomically validated the reconstructions created. Seven enhanced CT-project files were then automatically fitted, separately in different regions, to obtain a single dataset of superimposed BP regions that incorporated anatomical variation. From this dataset, improved BP contouring guidelines were developed. These guidelines were then applied to the seven original CT-project files, and also to one additional file, left out from the superimposing procedure. The percentage of BP inclusion was compared with the published guidelines.

Results: The anatomical validation procedure showed a high level of conformity for the BP regions examined between the 3D-reconstructions generated and the dissected counterparts. Accurate and detailed BP contouring guidelines were developed, which provided a corresponding guidance for each level in a clinical dataset. An average margin of 4.7 mm around the anatomically validated BP contour is sufficient to accommodate for anatomical variation. Using the new guidelines, 100% inclusion of the BP was achieved, compared to a mean inclusion of 37.75% when published guidelines were applied.

Conclusion: Improved guidelines for BP delineation were developed using combined MRI and CT imaging with validation by anatomical dissection.

Introduction

The complete change from the use of a traditional radiation field design based mostly on bony landmarks to that of anatomical contouring of the tumor and organs at risk (OAR), followed by the use of computer algorithms to optimize dose distributions by means of inverse planning, called for a detailed knowledge of tumor as well as normal tissue anatomy and their variations (1). Moreover, imprecise delineations continue to be the key obstacle for achieving high geometric accuracy in IMRT, especially for complex anatomical regions (2). Imprecise delineations represent a potential

source of uncertainties in historical dose/volume/effect data, thereby leading to reduced performance of predictive models (2).

One particularly complex anatomical region hosts the BP, which passes through the interscalene, costoclavicular, and subcoracoid spaces from the lower cervical spine to the upper limb. Often, the BP is also adjacent to –or may be invaded by- the primary tumor, pathological nodes, and elective nodal volumes, and therefore, needs to be carefully delineated (1). Inaccurate or incomplete contouring of the BP may leave branches insufficiently constrained and may result in dose distributions that exceed tolerance of these branches (3).

According to recent clinical investigations, up to 22% of head-and-neck cancer patients treated with radiation therapy developed upper extremity symptoms, which were consistent with BP injury and may be related to radiation dose (4).

Existing delineation guidelines for the BP (1,3,5) are imprecise, incomplete, and have not been anatomically validated. For example, contouring guidelines for regions of the BP lateral to the interscalene triangle are still referred to as being at a fixed vertebral level (3,5). This recommendation is untenable based on the degree of variation in this region, increasing with distance from the vertebral column. This is primarily due to the highly mobile claviculoscapular complex that provides a bony interface to determine the course of the BP. Furthermore, the existing guidelines are elaborated only for a limited number of slice levels (1,3,5). Therefore, when non-indicated slice levels need to be delineated, key information is lacking. Additional details are not clearly defined, and thus are subject to interpretation. For example, should the T1-T2 intervertebral foramen be delineated (3)? Moreover, the use of a variety of terms for anatomical orientation (3) and the designation of the BP as a reference point (1) can lead to ambiguity. As a result, it is difficult for clinicians to achieve a high-quality contour delineation of the BP.

Finally, the validation procedure for a contoured structure according to existing guidelines is based on expert opinion and consensus (1,3,5). Moreover, until now, the concept of anatomical validation has never been used for the development of delineation guidelines. Most parts of the BP are not directly visible on imaging for radiotherapy planning. Therefore, the omission of an anatomical validation process can result in a significant underestimation of the normal anatomical variation of the BP and its surroundings, as well as a higher risk of error (6). Therefore, the aim of this study was to develop contouring guidelines for the BP based on anatomically validated datasets obtained using a fusion of high resolution magnetic resonance imaging (MRI) and computed tomography (CT).

Methods and materials

Eight cadavers (1 male, 7 female; mean age: 73 years) were embalmed according to the Thiel-method (7), and were used for medical imaging (CT, MRI) and dissection. Thiel-cadavers are characterized by flexible soft tissues and joints, full range of motion, and maintenance of tissue integrity (8). Furthermore, the use of Thiel-cadavers has previously provided very accurate models, resulting in high quality data for biomechanical and dosimetrical studies (8,9,10). In addition, Thiel-cadavers provide the opportunity to tolerate long MRI scanning times, unlimited radiation exposure (for CT), optimal standardization of body posture between CT and MRI, and to exclude movement, flow, and breathing artifacts.

Cadavers were positioned on a carbon plate in supine with neutral cervical spine and shoulder position and both arms alongside the body, previously referred to as an IMRT treatment position (11). The head, neck, and shoulders were immobilized using thermoplastic material (Orfit®) and straps. First, a MRI dataset was obtained for the right side using a 3T high-field TIM TRIO MRI-scanner (Siemens®) with a T1-weighted fat suppressed VIBE sequence (voxel 0.83 × 0.83 × 0.83 mm), 320 mm field-of-view, 384 × 384 pixel, 0.8 mm slice thickness, and a scanning time of 32 min (12). The cadaver was then transported on the carbon plate, without removing the fixation material, to the IMRT planning facility. There, the right side was analyzed using a helical CT-scan (Toshiba® Aquilion) with the following scanning parameters: 120 kVp, 300 mAs, slice increment 1 mm, 502.78 mm field-of-view, and 512 × 512 pixels (0.982 mm pixel size).

CT and MRI image datasets were separately imported into the three-dimensional (3D) software package, Mimics 15.0®, resulting in separate CT-project and MRI-project files. The original MRI dataset was combined with the CT dataset using ‘image registration’ (Mimics), in order to obtain the transformation matrix that was necessary for further processing. In the MRI-project files, soft tissue reference structures, which were necessary for guideline development, were semi-automatically segmented and 3D reconstructed. In contrast, the BP required a (non-automatic) manual segmentation procedure (12). Bony structures were reconstructed from CT-project files. All 3D-objects from the MRI-project file were imported into the CT-project file as stereolithographs (STLs), and were appropriately positioned by applying the transformation matrix of the initial image fusion to generate an 'enhanced CT-project' file.

The right cervical-shoulder region of each cadaver was then dissected according to an in-house protocol developed to preserve local topography (12). The 3D reconstructions for the BP and surrounding regions were compared with the corresponding regions in the dissected cadaver, and this verification resulted in a final ‘operational CT-project’ file.

To provide a general integration of anatomical variation, six individual BP regions were superimposed onto one master operational CT-project file using local coordinate transformations (translation-rotation-scaling) as defined by the respective morphology and position of proximal anatomical reference landmarks. This transformation procedure was driven by a stepwise integrated Matlab® code using a point correspondence and procrustes function. One operational CT-project was omitted from the protocol to perform an extra validation experiment.

This superimposing procedure was repeated for four consecutive independent regions from mediosagittal to lateral since the topography of the entire BP cannot be related. For example, if only the vertebral columns are superimposed, the location of the BP will be accurate according to its medial regions, which are close to the vertebral foramina, yet will significantly vary in its lateral regions due to inter-subject variation.

In the first independent region (vertebral region), vertebral bodies were superimposed for each cervical level. In the second region (scalene outlet), superimposition of the scalene muscles was performed. In the third region (sublavius-serratus outlet), the subclavian, serratus anterior, anterior scalene, and minor pectoral muscles were fitted due to their close relationship with the BP. Lastly, in the fourth region (the muscle triangle), the axillary artery, minor pectoral, subclavian, serratus anterior, coracobrachial muscle, and the second rib were fitted (Fig. 1).

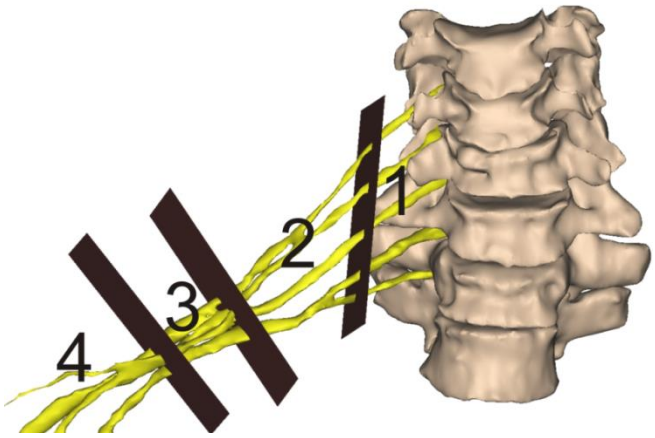


Fig. 1. The course of the BP was divided into four anatomical regions in order to independently superimpose these regions relative to proximal reference structures. 1, vertebral region; 2, scalene outlet; 3, subclavius-serratus outlet; 4, muscle triangle.

Following this 3D fitting procedure, the overlaid plexuses and surrounding regions were visualized in axial slices as a cross-sectional clustering of superimposed contours (Fig. 2). Then, for each axial slice, a minimal, yet convenient, delineation area was established to encapsulate the superimposed brachial plexuses. The clusters of contours, corresponding to the specific reference structures, were also used to deduce the appropriate general guidelines for describing the

delineation process used to achieve suitable BP inclusion for each individual case. These general guidelines were tested on seven original CT-projects together with an additional validation according the 'leave-one-out' principle on the eighth dataset. Following this principle, one anatomically validated imaging dataset has been left out from the development of the main model from which the contouring guidelines are deduced. This dataset was used for an extra validation experiment. Finally, the recommended guidelines of the current study and the existing guidelines (3) were applied to each CT-project separately, followed by a 'wrap function' (Mimics), to create 3D visualizations of the delineation volumes calculated. The percentage of BP inclusion, as the intersection volume divided by the anatomically validated BP volume, was also measured for both guidelines. Also each delineated volume was compared with the actual BP volume, and the average distance from the anatomically validated BP contour margin to the delineation margin was calculated.

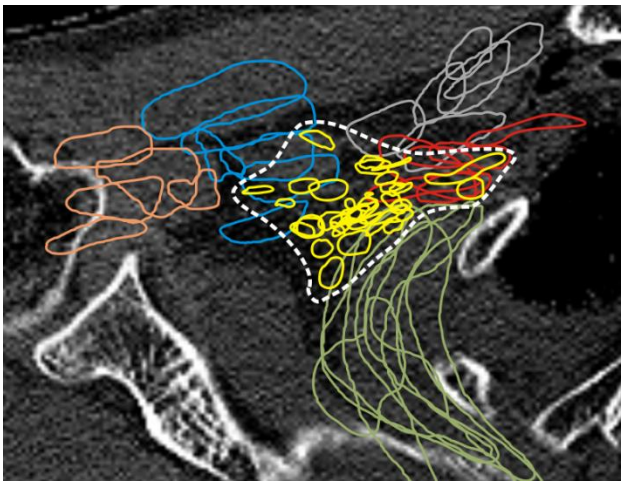


Fig. 2. Two-dimensional axial section through the 3D superimposed structures of six specimens. Clusters of superimposed muscles (gray, subclavian muscle; blue, minor pectoral; orange, coracobrachial muscle; green, serratus anterior) and axillary arteries (red) are visible. The BP regions are superimposed (yellow). The delineation area is identified as a dotted line.

Results

When 3D reconstructions of the BP and surrounding tissues were compared with the dissection, a high level of structural conformity was observed. For example, the branching and bifurcation patterns of the roots, trunks, divisions, cords, and peripheral nerves were identical to their genuine anatomical appearance. Furthermore, even anatomical variations were represented in the 3D reconstructions generated. Taken together, these results indicate that accurate reconstructions of anatomical regions can be achieved (Fig. 3).

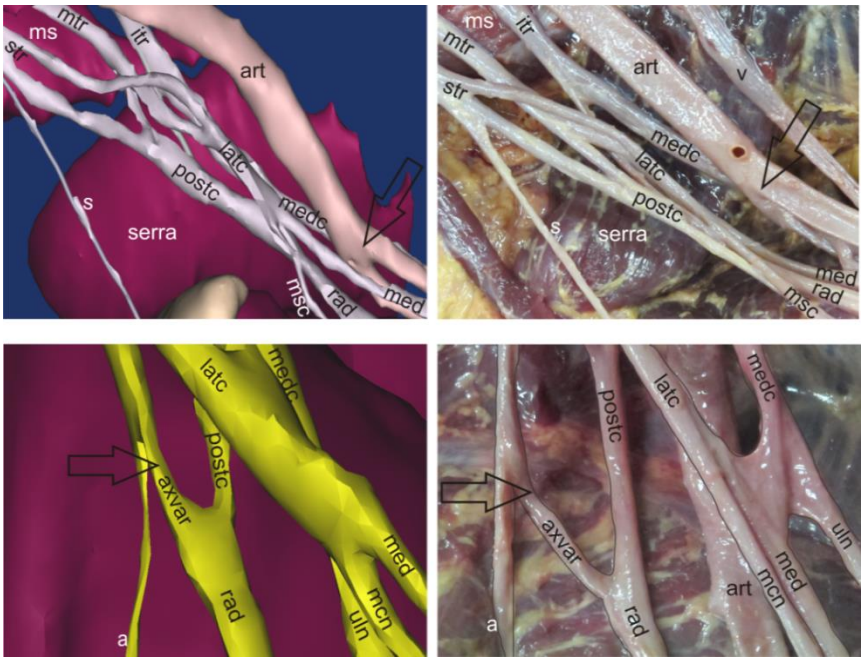
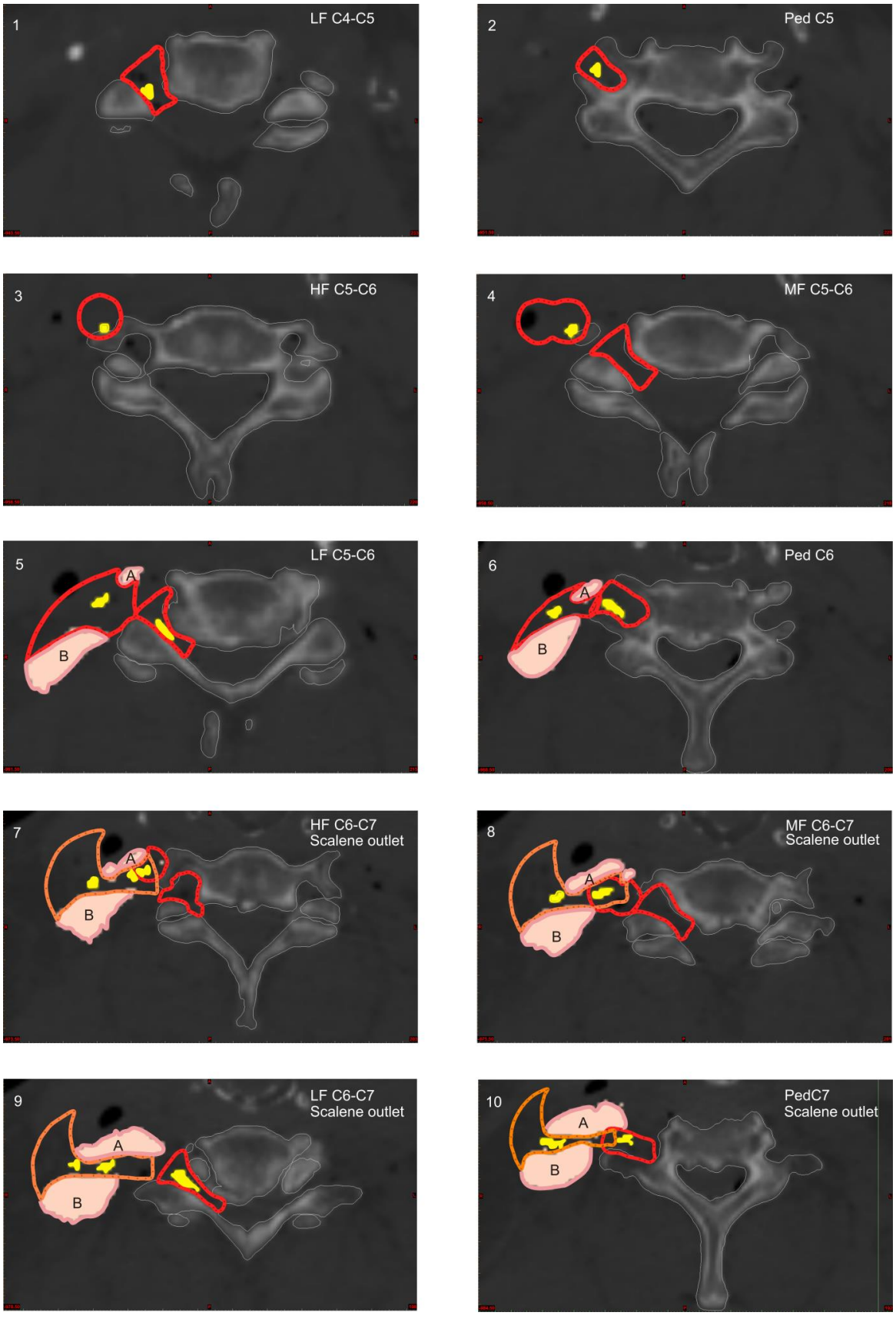


Fig. 3. A 3D-reconstruction of a BP is anatomically validated using cadaver dissection. Plexus details of two different cadavers are presented. Left top and bottom panels: reconstructions. Right top and bottom panels: dissections. In the first cadaver (upper panels), a bifurcation of the brachial artery is visible in both reconstruction and dissection views (open arrow). In the second cadaver (lower panels), an accessory axial nerve (axvar) connected with the posterior cord is present in both reconstruction and dissection views (open arrow). The axillary artery was not reconstructed, and the ulnar nerve is pulled medially in the dissection view.
str, superior trunk; mtr, middle trunk; itr, inferior trunk; postc, posterior cord; medc, medial cord; latc, lateral cord; ms, medial scalene; serra, serratus anterior; art, axillary artery; v, axillary vein; a, axillary nerve; rad, radial nerve; uln, ulnar nerve; med, median nerve; msc, musculocutaneous nerve; s, suprascapular nerve.

After transforming and superimposing seven different brachial plexuses and surroundings, the axial slices clearly illustrate the cross-sectional clustering of superimposed contours representing the reference structures (Fig. 2). The clusters show the spatial variation for the plexuses and each group of reference structures. In Figure 2, the white dotted line identifies the minimal delineation area that was generated, and which includes the superimposed brachial plexuses. It is this delineation area that constituted the starting point for developing guidelines in relation to clustered reference structures.

The guidelines identified in the present study are available at www.redjournal.org. In Figure 4, a detailed overview is provided of all the delineation levels identified for the various independent regions. Furthermore, the guidelines are presented in such a way that for each level in a clinical dataset, a corresponding guideline exists. Moreover, 1 mm increments were used, whereas treatment planning datasets tend to use 3 mm or 5 mm increments.


1
LF C4-C5
2
Ped C5
3
HF C5-C6
4
MF C5-C6
5
LF C5-C6
A
B
6
Ped C6
A
B
7
HF C6-C7
Scalene outlet
A
B
8
MF C6-C7
Scalene outlet
A
B
9
LF C6-C7
Scalene outlet
A
B
10
PedC7
Scalene outlet
A
B

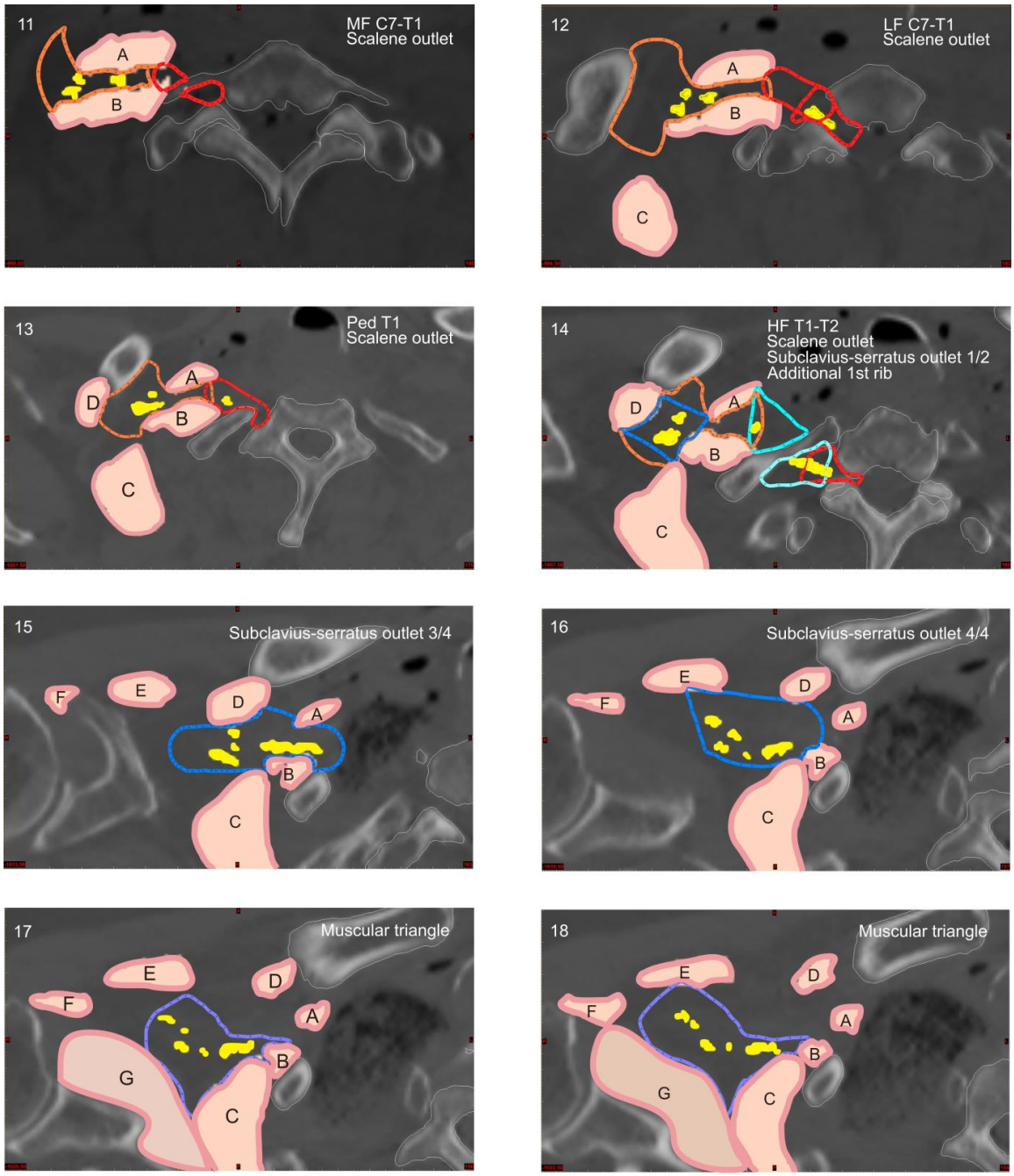


Fig. 4. Axial CT-sections for every delineation level for one cadaver. **A**: Delineation levels from Low Facet C4-C5 up to Pedicular C7 are shown and labeled sequentially, 1-10. **B**: Delineation levels from Mid Facet C7-T1 up to the muscular triangle (most lateral) are shown and labeled sequentially, 11-18. The BP contouring area is shown for each independent region: red (vertebral region), yellow (BP), pink (muscles), orange (scalene outlet), blue (subclavius-serratus outlet), green (additional first rib delineation), and purple (muscular triangle). The total delineation area at each level represents a summation of the delineations for each region. HF, high facet level; MF, mid facet level; LF, low facet level; Ped, pedicular level. Labels: A, anterior scalene; B, middle scalene; C, serratus anterior; D, subclavian muscle; E, minor pectoral; F, coracobrachial muscle; G, subscapular muscle.

The guidelines developed in the present study provided 100% inclusion for the BP region in the seven original datasets, as well as for the dataset that was left out. In contrast, the guidelines according to Hall et al. (3), which were subjected to an identical wrap function, resulted in a mean inclusion of 37.75% (Table 1, Fig. 5). Furthermore, the delineation volume achieved using the new guidelines was minimal, while still taking into account anatomical variations of the BP region (Table 1, Fig. 5). The average distance from the anatomically validated BP contour margin to the delineation margin was 4.7 mm.

Table 1. Brachial plexus coverage

Cadaver number	Covered plexus volume (%); recommended	Covered plexus volume (%); conventional	$V_{guidelines}/V_{plexus}$
1	100	25	4.55
2	100	50	3.38
3	100	33	4.07
4	100	42	4.62
5	100	50	3.29
6	100	38	3.87
7	100	32	3.40
8	100	32	3.65
Mean	**100**	**37.75**	**3.85**

Abbreviations: $V_{guidelines}$ = delineation volume by application of recommended guidelines; V_{plexus} = real plexus volume.

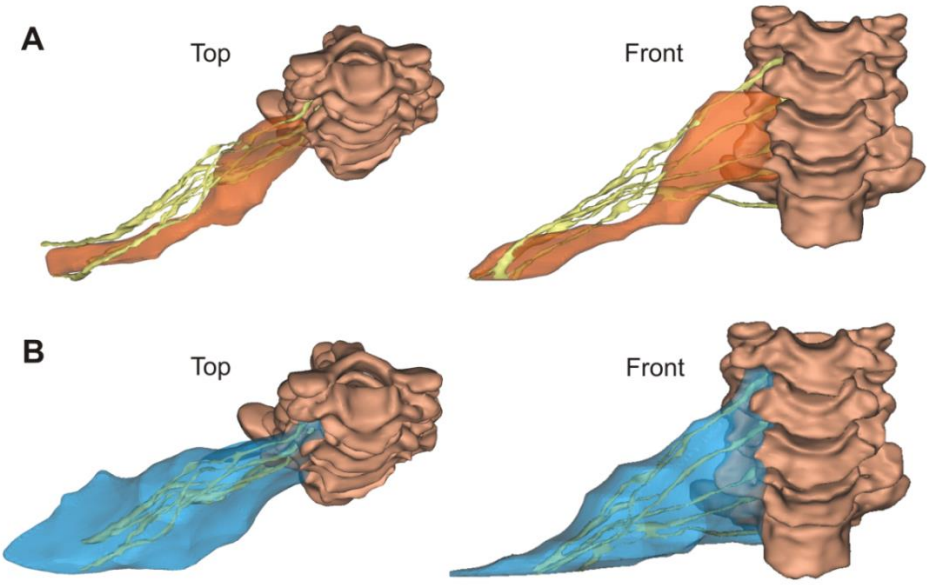


Fig. 5. 3D-contouring of the BP region: (**A**) conventional guidelines (orange, representing 37.75% plexus inclusion) and (**B**) the new guidelines of the present study (blue, representing 100% plexus inclusion). Left panels: top views, right panels: front views.

The reconstructions of the delineated volumes (wraps) also have a small area surrounding the roots, while the divisions are surrounded by a remarkably larger area with a mean increase of 25.34%. This serves to represent the dissimilarity of the anatomical variations that can exist between different regions of the BP (Fig. 5B).

Discussion

The main goal of this study was to develop refined guidelines for the accurate delineation of the BP using anatomically validated CT and MRI datasets. To our knowledge, this is the first study to successfully generate detailed 3D-reconstructions of the entire BP using MRI and without the use of any contrast agent (12). Existing 3D-reconstructions only provide detailed root depictions, while the details of distal regions of the BP are not shown (13-17).

Based on the high level of structural conformity achieved for the BP and surrounding regions in the virtual projections versus the dissected regions, the anatomical validity of the method used was demonstrated. Furthermore, the application of the new guidelines developed from the anatomically validated project files, resulted in 100% inclusion of the BP, which is a significant improvement over the 37.75% inclusion achieved using published guidelines.

There are several reasons for the significant difference in BP inclusion achieved. First, the new guidelines use unambiguous anatomical anchor points, thereby facilitating the identification of high quality delineations by clinicians. Also, more anatomical information is available using the new guidelines: a guideline exists for the clinical dataset associated with each slice. Moreover, the method of fitting structures separately in independent zones leads to accurate positioning of the major regions of the BP in relation to appropriate reference structures, while also accounting for natural anatomical variations.

According to the published guidelines, the majority of delineation areas, also the most exterior, are referred to a vertebral level. However, this can result in erroneous contouring due to an increase in positional variation that occurs with increasing distance from the vertebral column (3,5,18). For example, when level determination of the posterior aspect of the neurovascular bundle is related to the sternoclavicular joint, the aortic arch, and the glenohumeral joint, a greater incidence of delineation errors will occur (18). Moreover, published guidelines often rely on, and refer to, the visibility of the BP itself (1). This is very unreliable since neural structures can easily be confused with fat, connective tissue, or vessels.

The fusion of CT-MRI datasets for the published guidelines involved the following considerations. First, these fusions were performed following an indirect (different individuals) procedure and patient positioning was not standardized. In contrast, the current study used CT-MRI data obtained from the same individual (direct) and these images were obtained using a standardized body position. Secondly, the literature only reports expert opinion and consensus reading as tools to validate structure contouring (1,3,18). However, this is an interpretation-based approach, and does not accommodate anatomical variation. In contrast, the current study describes a method that provides the exact location of a subject's BP and incorporates anatomical variation (Fig. 4). In the clinic, the application of the new guidelines may lead to a more precise delineation of the BP, thereby minimizing the risk for brachial plexopathy (3,18). The average distance from the anatomically validated BP contour margin to the delineation margin was 4.7 mm. This margin incorporates delineation uncertainty and anatomical variation which are often disregarded in contouring guidelines.

Another advantage of the new guidelines is that delineation can be performed without prior knowledge of BP topography. This may save a considerable amount of time that would otherwise be spent scrutinizing anatomical and radiological atlases. Besides, the new guidelines can also be applied to the axillary portion of the BP which may be relevant in breast cancer and apical lung tumor treatment.. Further optimization of the guidelines is necessary for other treatment positions, such as prone or as positions with elevated arms. The new guidelines also make use of geometrical shapes such as lines and circles to identify the delineation area, and this is a straightforward task using

Mimics software. However, delineation software such as Pinnacle® is not equipped to drag and drop circles that are necessary for the current delineation. This can be addressed by either programming a circle drawing tool, or by delineating the area of interest using Mimics software, then importing the resulting areas as a DICOM structure into Pinnacle.

Conclusion

Brachial plexopathy is a complication of head-and-neck cancer treatment in which radiation therapy may play a role. Most parts of the BP are not directly visible on imaging used for radiation therapy planning. Published guidelines use surrounding structures to delineate a volume that is expected to encompass the BP. These guidelines lack anatomical validation. New guidelines for BP delineation were developed using combined MRI and CT imaging with validation by anatomical dissection using Thiel-embalmed cadavers. Anatomical variations of the BP and its surroundings are taken into account. The new guidelines provided 100% inclusion of the BP in all specimens, while minimizing the delineation volume. Further investigations aim at increasing the robustness of new guidelines using additional Thiel-specimens and at evaluating the value of Thiel-specimen based atlases for autocontouring of the BP.

References

1. Truong MT, Nadgir RN, Hirsch AE, et al. Brachial plexus contouring with ct and mr imaging in radiation therapy planning for head and neck cancer. *Radiographics* 2010;30:1095-1103.
2. Brouwer CL, Steenbakkers RJ, Van den Heuvel E, et al. 3d variation in delineation of head and neck organs at risk. *Radiat. Oncol.* 2012;7.
3. Hall WH, Guiou M, Lee NY, et al. Development and validation of a standardized method for contouring the brachial plexus: Preliminary dosimetric analysis among patients treated with imrt for head-and-neck cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2008;72:1362-1367.
4. Lundstedt D, Gustafsson M, Steineck G, et al. Long-term symptoms after radiotherapy of supraclavicular lymph nodes in breast cancer patients. *Radiother. Oncol.* 2012;103:155-160.
5. Kong FM, Ritter T, Quint DJ, et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *Int. J. Radiat. Oncol. Biol. Phys.* 2011;81:1442-1457.
6. Kerr AT. The brachial plexus of nerves in man, the variations in its formation and branches. *Am. J. Anat.* 1918;23:285-395.
7. Thiel W. Die Konservierung ganzer leichen in natürlichen farben. *Ann. Anat.* 1992; 174: 185-195.
8. De Crop A, Bacher K, Van Hoof T, et al. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. *Radiology* 2012;262:298-304.
9. Audenaert EA, Vigneron L, Van Hoof T, et al. In vitro validation and reliability study of electromagnetic skin sensors for evaluation of end range of motion positions of the hip. *Med. Biol. Eng. Comput.* 2011; 49:1405–1412.
10. Schuenke MD, Vleeming A, Van Hoof T, et al. A description of the lumbar interfascial triangle and its relation with the lateral raphe: anatomical constituents of load transfer through the lateral margin of the thoracolumbar fascia. *J. Anat.* 2012; 221:568–576.
11. Neubauer E, Dong L, Followill DS, et al. Assessment of shoulder position variation and its impact on imrt and vmat doses for head and neck cancer. *Radiat. Oncol.* 2012;7.
12. Van Hoof T, Gomes GT, Audenaert E, et al. 3d computerized model for measuring strain and displacement of the brachial plexus following placement of reverse shoulder prosthesis. *Anat. Rec.* 2008;291:1173-1185.
13. van Es HW, Bollen TL van Heesewijk HPM. Mri of the brachial plexus: A pictorial review. *Eur. J. Radiol.* 2010;74:391-402.

14. Tagliafico A, Succio G, Neumaier CE, et al. Brachial plexus assessment with three-dimensional isotropic resolution fast spin echo mri: Comparison with conventional mri at 3.0 t. *Br. J. Radiol.* 2012;85:E110-E116.
15. Vargas MI, Viallon M, Nguyen D, et al. New approaches in imaging of the brachial plexus. *Eur. J. Radiol.* 2010;74:403-410.
16. Chen ZG, Zang J, Chen T, et al. Study of sensory and motor fascicles in brachial plexus and establishment of a digital three-dimensional graphic model. *Ann. Plast. Surg.* 2011;67:615-619.
17. Kjelstrup T, Courivaud F, Klaastad Ø, et al. High-resolution mri demonstrates detailed anatomy of the axillary brachial plexus. A pilot study. *Acta Anaesthesiol. Scand.* 2012;56:914-919.
18. Yi SK, Hall WH, Mathai M, et al. Validating the rtog-endorsed brachial plexus contouring atlas: An evaluation of reproducibility among patients treated by intensity-modulated radiotherapy for head-and-neck cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2012;82:1060-1064.

Reply letter to the editor:

An anatomically validated brachial plexus contouring method for intensity modulated radiation therapy planning: reply.

Joris Van de Velde, Wilfried De Neve, Tom Van Hoof

Int J Radiat Oncol Biol Phys 2014;89:224-225. Published

IF: 4.176.

COMMENTS:

An anatomically validated brachial plexus contouring method for intensity modulated radiation therapy planning

In Reply to Basu et al.

We would like to thank Basu et al. for their comments on our publication 'An Anatomically Validated brachial plexus contouring method for intensity modulated radiation therapy planning' and we would like to address the issues raised.

We agree that brachial plexopathy is uncommon in Head-and-neck Cancer (HNC) IMRT although it may be underreported. Chen et al (2011) recently reported an incidence of 13% of neuropathic symptoms after radiation therapy for HNC cancer. When patients with less than 5 years of follow-up were excluded, the rate of positive symptoms increased up to 22% (1).

In regard to the comment that we did not mention properly whether the contours were fused to make a common volume after applying the guidelines at each level, the authors can state that this was clearly described in the first paragraph of the second column of page 804: "Finally, the recommended guidelines of the current study and the existing guidelines were applied to each CT project separately, followed by a wrap function to create 3D visualizations of the delineation volumes calculated"(2).

The commentary states that major concern for brachial plexopathy arises when a need occurs to treat the lower neck at high doses in a post-operative or level IV node-positive neck. This statement is consistent with brachial plexopathy observed in breast cancer patients after supraclavicular-axillary lymph node irradiation, apparently occurring at lower doses than the 60-66 Gy, reported to be safe higher up in the neck. A recent review from Delanian et al. showed up to 12% brachial plexopathy after mastectomy + axillary dissection levels I-II when using 50 Gy in 25 fractions (3). It must be emphasized that the length of irradiated brachial plexus is higher in breast cancer than in HNC, especially when all axillary node levels are included. Large fraction size and technical flaws (overlapping fields, daily alternating-field treatment) strongly influenced the incidence of brachial plexopathy. The observation regarding fraction size is relevant for HNC, considering modern tendencies to apply simultaneous integrated boost, dose-painting or stereotactic boost techniques.

The fraction/dose/length/toxicity relationship using the newly delineated brachial plexus volumes is unknown. The question to conduct a clinical correlation study is well taken. We indeed plan such study.

1. Chen AM, Hall WH, Li J, et al. Brachial plexus-associated neuropathy after high-dose radiation therapy for head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2012;84(1):165-9.

2. Van de Velde J, Audenaert E, Speleers B, et.al. An Anatomically Validated Brachial Plexus Contouring Method for Intensity Modulated Radiation Therapy Planning. *Int J Radiat Oncol Biol Phys* 2013;87:802-808.

3. Delanian S, Lefaix JL, Pradat PF. Radiation-induced neuropathy in cancer survivors. *Radiother Oncol* 2012;105(3):273-82. (review)

Manuscript 4

Morphometric atlas selection for automatic brachial plexus segmentation

Joris Van de Velde, Johan Wouters*, Tom Vercauteren, Werner De Gersem, Frédéric Duprez, Wilfried De Neve, Tom Van Hoof

Int J Radiat Oncol Biol Phys 2015. Published

A1, IF: 4.176. *JVdV and JW are shared first authors

Abstract

Purpose*:* The purpose of this study was to determine the effects of atlas selection, based on different morphometric parameters, on the accuracy of automatic brachial plexus (BP) segmentation for radiotherapy planning. The segmentation accuracy was measured by comparing all of the generated automatic segmentations with anatomically validated gold standard atlases developed using cadavers.

Methods and materials: Twelve cadaver computed tomography (CT) atlases (3 males, 9 females; mean age: 73 years) were included in the study. One atlas was selected to serve as a patient, and the other 11 atlases were registered separately onto this 'patient' using deformable image registration provided by the ABAS®1.1 software. This procedure was repeated for every atlas as a patient. Next, the Dice and Jaccard Similarity Indices and Inclusion Index were calculated for every registered BP with the original gold standard BP. In parallel, differences in several morphometric parameters that may influence the BP segmentation accuracy were measured for the different atlases. Specific brachial plexus-related, CT-visible bony points were used to define the morphometric parameters. Subsequently, correlations between the similarity indices and morphometric parameters were calculated.

Results: A clear negative correlation between difference in protraction-retraction distance and the similarity indices was observed [mean Pearson's correlation coefficient (PCC) = -0.546]. All of the other investigated PCC values were weak.

Conclusions: Differences in the shoulder protraction-retraction position between the atlas and the patient during planning CT influence the BP autosegmentation accuracy. A greater difference in the protraction-retraction distance between the atlas and the patient reduces the accuracy of the BP automatic segmentation result.

Introduction

The brachial plexus (BP) is an organ at risk in patients with lung, breast and head-and-neck cancer who undergo radiotherapy. Accurate BP segmentation on computed tomography (CT) is crucial to protect this organ against plexopathy after high-dose radiotherapy. Often, the BP is also adjacent to, or may be invaded by, the primary tumor, pathologic nodes, and elective nodal volumes and therefore needs to be carefully delineated (1). Inaccurate or incomplete contouring of the BP may leave branches insufficiently constrained and may result in dose distributions that exceed the

tolerance of these branches (2). According to recent clinical investigations, up to 22% of head and-neck cancer patients treated with radiation therapy experienced upper extremity symptoms, which were consistent with BP injury and may be related to radiation dose (3). Because manual contouring is difficult and time-consuming, automation of this process would stimulate contouring and planning dosimetry of this organ at risk. During the automation of a target image-contouring process, multiple a priori segmented image datasets, which are called atlases, are used as templates. In the past, a major problem with BP segmentation was the lack of validated 'gold standard' templates. Templates were based on expert opinions or a consensus reading (1,2,4). Because the BP is invisible on the CT, these methods are unreliable for the development of a gold standard. Therefore, Van de Velde et al. created an innovative anatomically validated method for development of BP gold standard segmentations using cadavers (6,7).

To achieve accurate automatic BP segmentation, not only are excellent gold standards needed but also the number of atlases used must be sufficient (8). When an optimal number of atlases is achieved, no further improvement of segmentation will occur. At that stage, the inclusion of additional atlases will only expand the segmentation zone, resulting in an increase in the false positive delineation area (9). At that moment, further improvement can only be achieved by selecting the most corresponding atlases in the database.

Several methods for selecting corresponding atlases have been published (8- 10). Most of these methods are based on the similarity between the atlas and the target image. Atlas selection strategies based on image similarities has a number of drawbacks. When an insufficient similarity exists between the atlas and the target image, local mismatches often occur, which in turn leads to segmentation errors (11, 12, 13).

Within similarity-based atlas selection strategies, a further distinction can be made between methods that use registration of the whole image (8) and those that use only part of the image (14, 15). When only a part of the image is used, it remains questionable which part should be selected. Theoretically, the best option is to use only image regions that are topographically stable in relation to the structure that has to be segmented (16).

An alternative technique of atlas selection is through the use of meta-information related to the subject. The potential selection criteria are the body mass index, age, pathology, clinical history, gender and handedness (9). For example, age-based selection has been shown to be as effective as selection based on image similarity after affine transformation (17). One major disadvantage of using meta-information is that it is not suitable when dealing with the anatomical variability that occurs independently of the simple meta-information (9).

The weakness of image similarity selection methods that utilize largely irrelevant areas of anatomical information and also the weakness of meta-information selection methods that do not address the anatomical variability could both be countered through the use of only target-specific stable anatomical information in the selection process. However, a target-specific atlas selection method for automatic BP segmentation has never been investigated. To develop an effective atlas selection strategy for specific BP autosegmentation, it is crucial to define stable CT-measurable morphometric parameters that determine the shape of the organ. These parameters can be related both to inter-individual anatomical variance and to variance in the positioning of the patient. In the current study, a number of morphometric parameters, which were measured using stable bony structures, were investigated based on their influence on atlas-based BP autosegmentation accuracy.

The purpose of this study was to measure the effect of atlas selection, based on different morphometric parameters, on automatic BP segmentation accuracy.

Materials and methods

Gold standard development

For the development of gold standard atlases for BP contouring, Thiel-embalmed cadavers were used because of their optimal image quality and movement capacities; the latter allowed the required standardization of the scan position (18, 19). Magnetic resonance imaging (MRI) [T1-weighted fat-suppressed VIBE sequence (0.83 × 0.83 × 0.83 mm voxel, 320 mm field-of-view, 384 × 384 pixel, 0.8-mm slice thickness and 32-min scan time)] of the head-and-neck region was obtained to generate high-quality BP delineations that were anatomically validated by dissection (Fig.1). These anatomically validated MRI BP delineations were then rigidly fused to the corresponding CT (scanning parameters: 120 kVp, 300 mA, 1-mm slice increment, 502.78 mm field-of-view and 512 × 512 pixels) to obtain BP gold standard delineations that were applicable to the radiotherapy planning system. A detailed description has been provided by Van de Velde et al. (2013) (6).

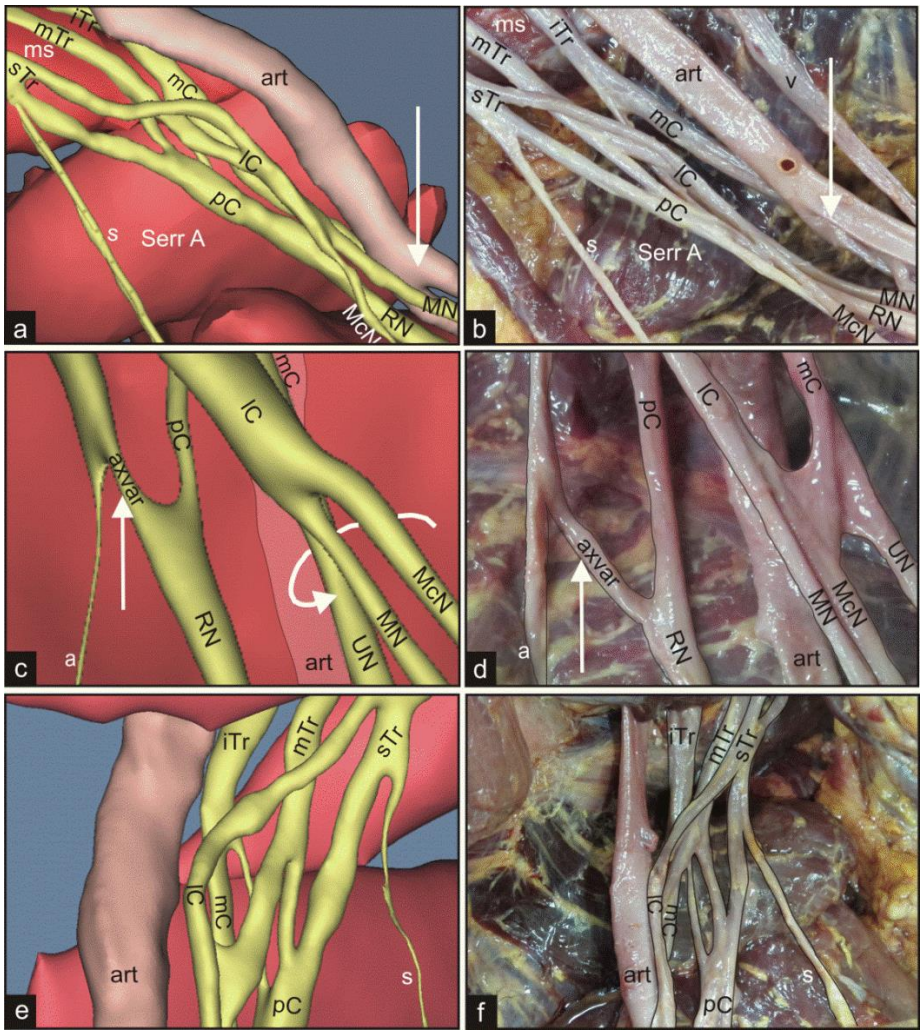


Fig. 1. 3D-reconstructions of brachial plexuses are anatomically validated by cadaver dissection. Brachial plexus details of three different cadavers are presented. In the first cadaver (a, b), a bifurcation of the brachial artery is visible in both reconstruction (a) and dissection views (b) (white arrows). In the second cadaver (c, d), an accessory axial nerve (axvar) connected with the posterior cord is present in both reconstruction and dissection views (white arrows). The axillary artery (art) was not reconstructed, the median and musculocutaneous nerves are rotated around the ulnar nerve due to abduction-exorotation position of the shoulder in the dissection view (curved arrow).

sTr, superior trunk; mTr, middle trunk; iTr, inferior trunk; pC, posterior cord; mC, medial cord; lC, lateral cord; ms, middle scalene; Serr A, serratus anterior; subsc, subscapular muscle; art, axillary artery; v, axillary vein; a, axillary nerve; RN, radial nerve; UN, ulnar nerve; MN, median nerve; McN, musculocutaneous nerve; s, suprascapular nerve.

The current study included 12 gold standard atlases (3 males, 9 females; mean age: 73 years), of which eight were also used in the study by Van de Velde et al. (2013). One atlas was selected to serve as a patient, and the other 11 atlases were registered separately onto the 'patient' using a deformable image registration ('General' algorithm) in ABAS®1.1 (Elekta AB, Stockholm, Sweden). The algorithm first aligns the atlas to the patient using a Dense Hybrid deformable image registration.

It consists of three major steps: a linear registration and two non-linear registration steps. Each step has increased degrees of freedom, and is used to provide initialization for the next step. The delineation of the atlas BP is then mapped to the patient using the transformation of this initial image registration to create the autosegmentation result (20). This procedure was repeated for every single atlas as a patient. This process resulted in 11 registered brachial plexuses onto every single atlas (n= 12) (11x12= 132 total deformations).

Then, for accuracy assessment of the autosegmentations, similarity indices between the registered BP and the original gold standard BP were calculated.

First, the Dice Similarity Coefficient (DSC) was calculated between the two segmentations (A, B). The DSC measures the spatial overlap between the gold standard A and the registered image B and is defined as DSC (A,B)= 2(A∩B)/(A+B), where ∩ is the intersection volume. The DSC is situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

We also calculated the Jaccard similarity Index (JI), as the ratio of the intersection volume and the entire union volume of the delineations JI(A,B)= (A∩B)/(AUB) . The JI is also situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement. The JI can be calculated from DSC with the formula [JI = DSC/(2-DSC)].

Finally, the Inclusion Index (INI) was measured between the gold standard BP (A) and the registered BP (B). INI is the intersection volume of both, divided by the gold standard BP INI=(A∩B/A). INI is situated between 0 and 1, with 0 indicating no inclusion and 1 indicating total inclusion of A by B (5).

Definition of morphometric parameters

To define morphometric parameters that potentially influence the segmentation quality, three neighboring bony structures, with a stable location in relation to the BP, were indicated on every CT-atlas. In all of the atlases, the BP generally had a triangle shape easily defined by three bony points: i) the anterior tubercle of the transverse process of C5, ii) the most anterior point of the head of the first rib and iii) the infraglenoid tubercle (the most caudal point of the glenoid fossa) (Fig. 2).

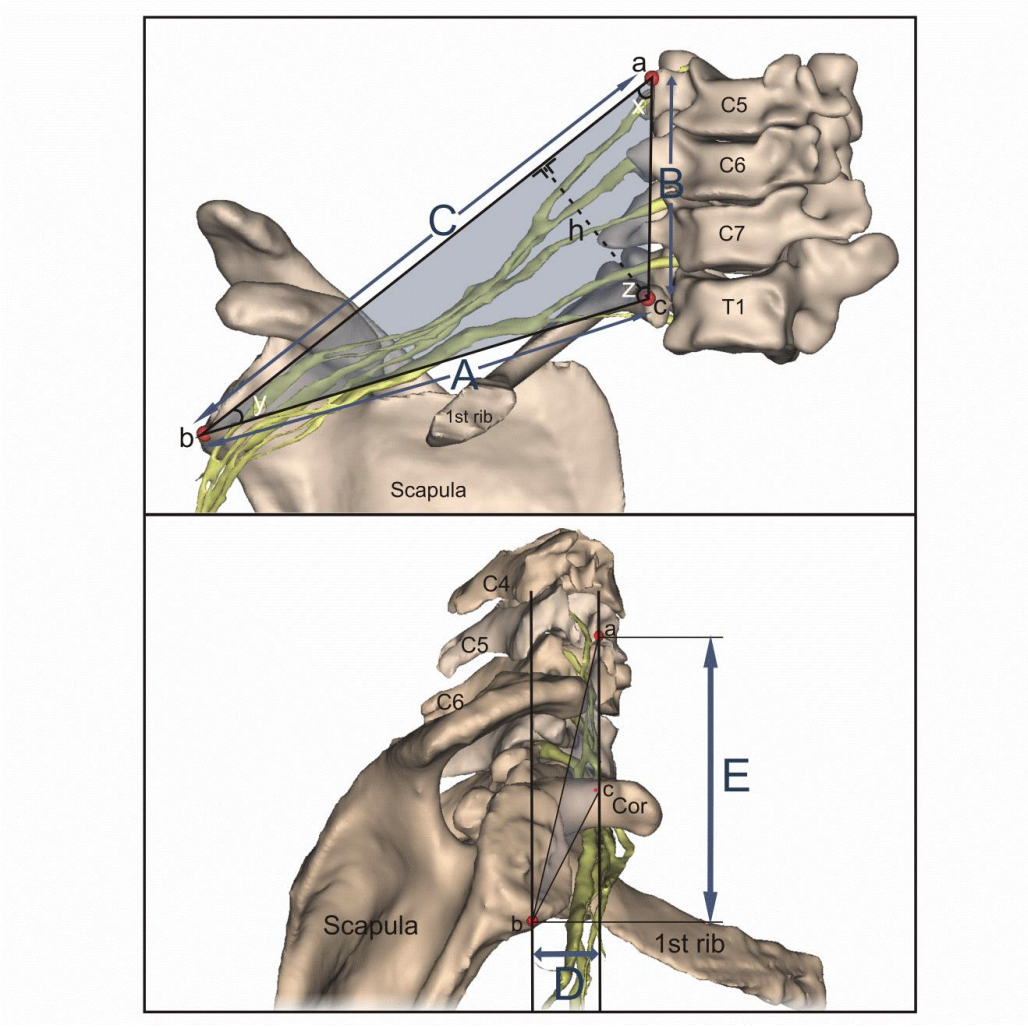


Fig. 2. Frontal (above) and lateral (below) view of a 3D illustration of the indication of the bony points near the brachial plexus (yellow) in a CT dataset. (a, anterior tubercle C5; b, infraglenoid tubercle; c, head of the first rib; D, protraction-retraction distance; E, elevation-depression distance; x, upper triangle angle; y, lateral triangle angle; z, lower triangle angle; h, height; Cor, coracoid process).

First, absolute distances (length) between these 3 points were calculated. Second, the surface (A x h/ 2) of the triangles were calculated (Fig. 2; A, B, C). The third parameter was the protraction-retraction distance ('protraction') (Fig.3), which was defined as the horizontal distance measured in the sagittal plane, between a vertical line through the anterior tubercle of C5 and a vertical line through the infraglenoid tubercle (Fig. 2; D). The fourth parameter was the lateralization parameter ('lateralization'), which was defined as the shoulder width/neck height ratio (Fig. 2; A/B). At last, the parameter elevation-depression ('elevation') (Fig.3) was measured. This is the vertical distance between the horizontal line through the anterior tubercle C5 and the horizontal line through the infraglenoid tubercle (Fig. 2; E).

Finally, for every pair, the difference between both atlases was calculated, concerning each individual morphometric parameter (Δ parameter).

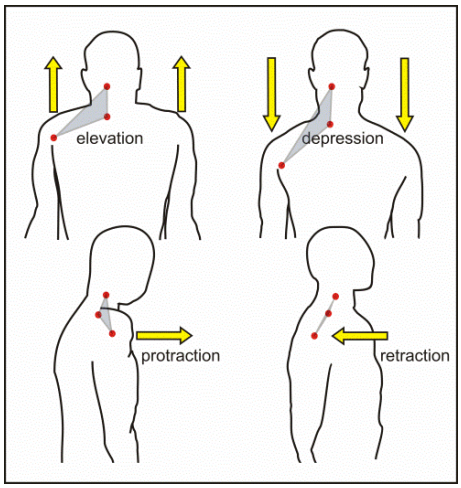


Fig. 3. Illustration of shoulder elevation-depression and protraction-retraction movement.

Statistical analysis

Pearson's correlation coefficients (PCC) were calculated between all the similarity measurements (DSC, JI, INI) and all the Δ parameters and this for every pair of atlases (Fig. 4). PCC is a measure of the linear correlation between two variables, providing a value between +1 and −1, where 1 is the total positive correlation, −1 the total negative correlation and 0 means no correlation. A labelling systems was used to roughly categorize PCC, where correlation coefficients that are < 0.35 are generally considered to represent weak correlations, 0.35 to 0.67 moderate correlations, and 0.68 to 1 high correlations (21).

To exclude disturbing influences of other parameters like the number of atlases, a single-atlas approach was chosen for this study. Because a single-atlas approach was applied, decreased similarity index values are expected relative to a multi-atlas approach.

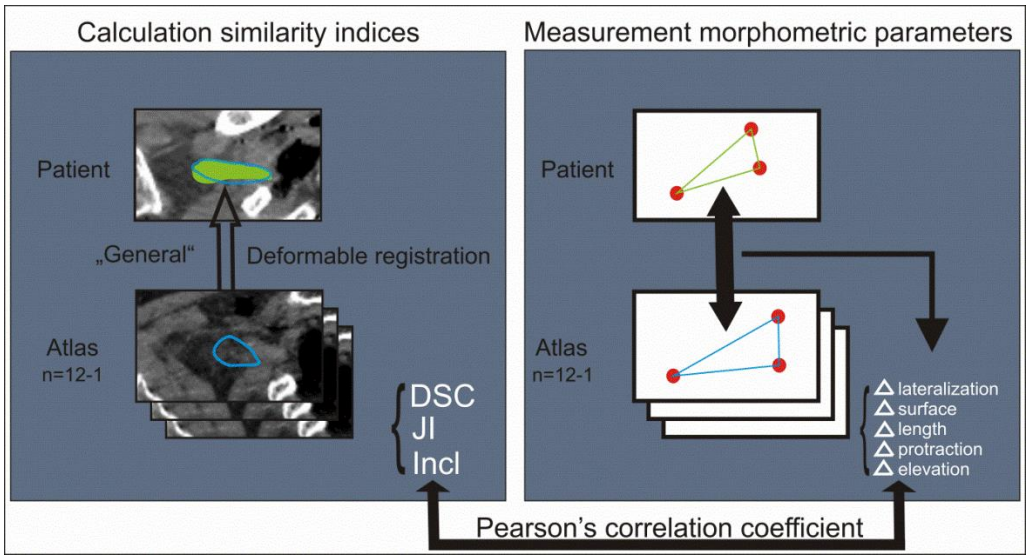


Fig. 4. Schematic illustration of the procedure used for one patient. This procedure was repeated for every atlas as a patient (n=12).

Results

The PCC between the similarity indices and the different morphometric parameters are shown in table 1. In addition, the mean absolute value and standard deviation for every Δ parameter were calculated.

Table 1. The mean absolute values for all of the Δ parameters and Pearson's correlation coefficients between the similarity indices and the different morphometric parameters ($p<0.001$).

Morphometric parameter	Mean (SD)	DSC	JI	INI
Δ lateralization	0.49 (0.32)	-0.124	-0.11	-0.128
Δ surface	715.6 (477.0) mm^2	0.193	0.191	0.141
Δ length A	15.3 (12.0) mm	-0.194	-0.185	-0.171
Δ length B	7.8 (5.1) mm	0.077	0.077	0.025
Δ length C	18.1 (15.5) mm	-0.228	-0.218	-0.22
Δ protraction	20.2 (13.4) mm	-0.581	-0.551	-0.508
Δ elevation	13.8 (9.8) mm	-0.088	-0.102	-0.107

Abbreviations: SD, standard deviation; DSC, Dice similarity coefficient; JI, Jaccard index; INI, inclusion index

All of the investigated correlation values were statistically significant ($p < 0.001$). The correlations between Δ protraction and the similarity indices were moderate and potentially clinically relevant (mean PCC = -0.547) (Fig. 5A, B, C). All of the other PCC values were weak.

As shown in the 6 scatter plots presented in figure 5, each time the same 132 deformable registrations (132 points; 11 registered brachial plexuses onto every single atlas) are depicted, once for Δ protraction (left) and once for Δ lateralization (right) correlated with the 3 different similarity indices. For Δ protraction, the scatter plots display a moderate correlation. It is noteworthy that the absence of low similarity indices was associated with low Δ protraction values in the scatter plots. However, in the scatter plots for the Δ lateralization parameter, many of the low similarity indices are associated with low Δ lateralization values, resulting in weak PCC values for this parameter.

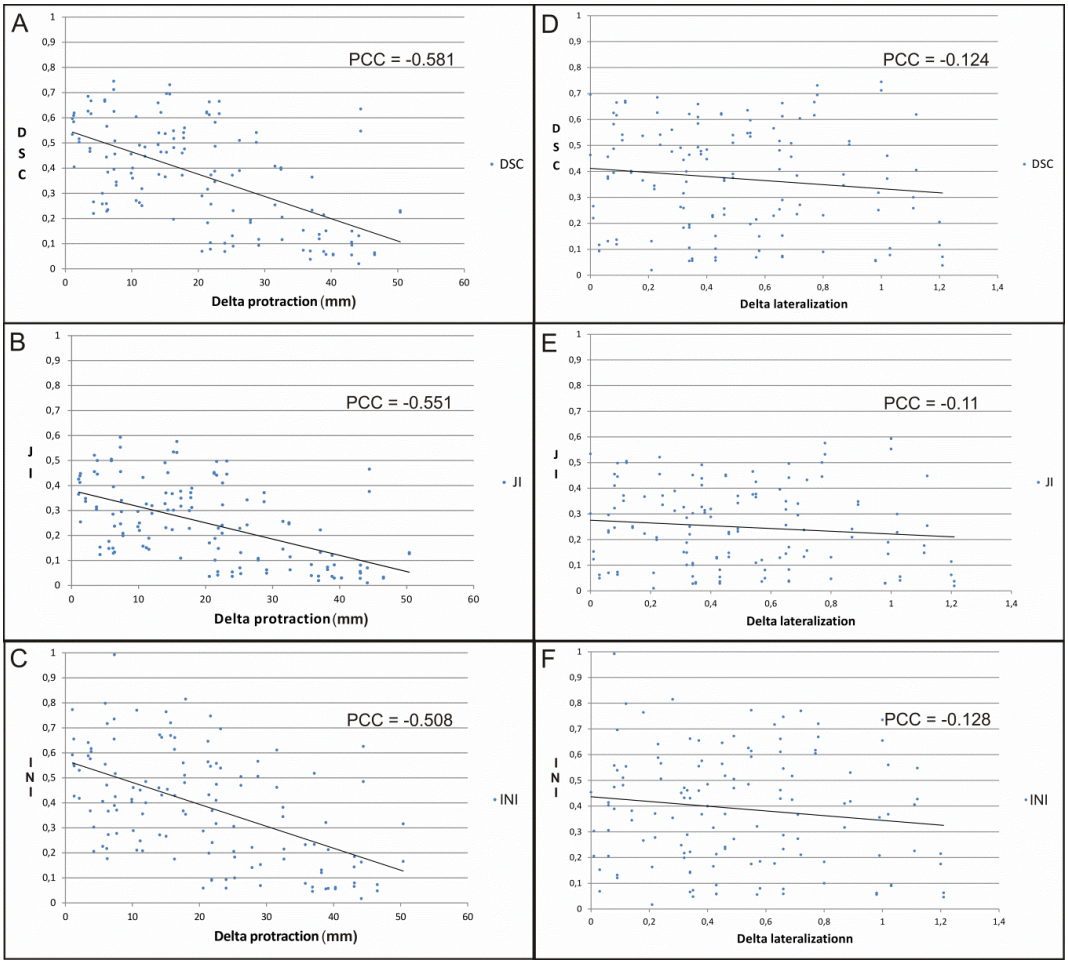


Fig. 5. Left**:** Scatter plots showing a clear negative correlation of the (A) Dice Similarity Coefficient, (B) Jaccard Index, (C) Inclusion Index and Δ protraction parameter. Right: Scatter plots showing a weak correlation of the (D) Dice Similarity Coefficient, (E) Jaccard Index, (F) Inclusion Index and Δ

lateralization parameter. DSC, Dice Similarity Coefficient; JI, Jaccard Index; INI, Inclusion Index; PCC, Pearson's Correlation Coefficient.

The Δ length A and Δ length B parameters correlations were very weak (Table 1). As expected, these absolute length differences had no influence on the segmentation accuracy. Surprisingly, Δ length C displayed a higher correlation than Δ lateralization (length A / length B), although it was not tested for significance. This could be explained by the fact that length C is defined by length A, B and the cosine of z (Fig. 2) ($C^2= B^2 +A^2- 2AB \cos z$). These three parameters appear to cause length C to become more sensitive to changes in BP morphology in comparison to Δ lateralization, which is only defined by the ratio of two parameters (length A and length B) with a low influence on the morphology of the triangle defining the shape of the BP.

In addition, a weak correlation was found for the Δ surface. The low PCC for Δ length A together with the observation that distance h (Fig. 2) did not define the shape of the BP might provide an explanation for the very weak correlation of the Δ surface with the similarity indices.

A comparison of the different parameters revealed that the PCC values had a parallel distribution (Table 1). Slight differences between similarity indices can be attributed to differences in the behavior of the similarity indices, with increases in the correctly delineated volume as shown in the supplementary material 1 (www.redjournal.org).

Discussion

The goal of this study was to measure the effect of morphometrically driven atlas selection on the accuracy of automatic BP segmentation.

For each investigated morphometric parameter (Δ lateralization, Δ surface, Δ length, Δ protraction and Δ elevation), the PCCs with similarity indices (DSC, JI and INI) showed statistical significance.

The most meaningful correlation was observed between Δ protraction and the similarity indices (mean PCC= -0.547 ($p<0,001$)) (Table 1, Figure 5A- C). This finding shows that when Δ protraction between the atlas and the patient dataset increased, the values for the similarity indices decreased, indicating inferior registration. Thus, greater differences in the protraction distance between the image datasets resulted in a reduced accuracy of the BP automatic segmentation results. These data suggest the extreme importance of a good scapular protraction positioning protocol of the patient to minimize the positional differences between the atlas and the patient.

For all of the other investigated parameters, a weak correlation with the similarity indices was found. Of note, there was no meaningful correlation for scapular elevation-depression. Strong effects of elevation-depression on the natural course -and thus also the shape of the BP- were expected based on previous anatomical reports and several biomechanical reports (7, 22-24). The observation that elevation-depression did not affect the BP autosegementation accuracy might be due to the applied methodology. We hypothesize that the lateral reference point b determining the elevation-depression parameter (Fig. 2; b) was also susceptible to scapular rotation due to its distance from the anteroposterior rotation axis, which is located just medial from the base of the coracoid process. Consequently, the scapular rotation will also alter the altitude of the reference point b and therefore affect the vertical distance E (Fig. 2), affecting the elevation-depression parameter. The base of the coracoid process, due to its proximity to the rotation axis, will show more robust responses to rotational movements of the scapula and is also a qualified determiner of the elevation-depression. This point, however, was not selected based on an inability to completely describe the BP in its most lateral extensions. A second reason to avoid selecting the coracoid process was that the dimensionality of its base complicates consistent determinations throughout the individual atlases.

The low correlations determined for Δ lateralization were also surprising. This phenomenon and the high Δ protraction correlations could be related to the distance between the BP and the clavicle in the antero-posterior direction. Because the clearest BP registration errors are observed in the subclavicular region, we assume that the anteroposterior BP-clavicular distance plays a crucial role in the registration process (supplementary material Fig. e2). The increased scapular protraction results in an increase in the antero-posterior distance between the BP and the clavicle (22), whereas alterations in lateralization would not or would barely result in a change in the BP-clavicle position in the anteroposterior direction. We hypothesize that a change in the position of the BP with respect to the clavicle in the antero-posterior direction would impede the ability of the algorithm to correctly register the BP in that area. The absence of a change in the position of the BP with respect to the clavicle will facilitate the ability of the algorithm to correctly register the BP.

As mentioned in the methodology section, low similarity indices were expected as a consequence of the single atlas approach. However, this phenomenon is irrelevant for the current study because only the effect of morphometric parameters on the registration accuracy was investigated. The overall development process for automated BP segmentation, of which the current study is an essential part, follows a stepwise approach. Additional atlas selection strategies will be investigated in the future, along with the optimal number of atlases and optimal label fusion algorithms for use in automatic BP segmentation. Finally, all of these optimized parameters will be applied in a multi-atlas autosegmentation protocol.

Conclusion

Correlation statistics demonstrate that differences in the shoulder protraction-retraction position between the atlas and the patient during planning CT influences the BP autosegmentation accuracy. A greater difference in the protraction-retraction distance results in a reduced accuracy of the BP automatic segmentation results and vice versa.

In the future, in existing automatic multi-atlas bases BP segmentation algorithms, we plan to implement an atlas selection strategy based on differences in protraction-retraction positions between the atlas and the patient. This, along with the optimization of other autosegmentation parameters, will potentially result in high quality automatic BP segmentations in the clinic.

References

1. Truong MT, Nadgir RN, Hirsch AE, et al. Brachial plexus contouring with ct and mr imaging in radiation therapy planning for head and neck cancer. *Radiographics* 2010;30:1095-1103.

2. Hall WH, Guiou M, Lee NY, et al. Development and validation of a standardized method for contouring the brachial plexus: Preliminary dosimetric analysis among patients treated with imrt for head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2008;72:1362-1367.

3. Lundstedt D, Gustafsson M, Steineck G, et al. Long-term symptoms after radiotherapy of supraclavicular lymph nodes in breast cancer patients. Radiother Oncol 2012;103:155-160.

4. Kong FM, Ritter T, Quint DJ, et al. Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *Int J Radiat Oncol Biol Phys* 2011;81:1442-1457.

5. Yang J, Amini A, Williamson R, et al. Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy. *PRO* 2013;3:139-147.

6. Van de Velde J, Audenaert EA, Speleers B et al. An anatomically validated brachial plexus contouring method for intensity modulated radiation therapy planning. *Int J Radiat Oncol Biol Phys* 2013;87:802-808.

7. Van Hoof T, Gomes GT, Audenaert E, et al. 3D computerized model for measuring strain and displacement of the brachial plexus following placement of reverse shoulder prosthesis. *Anat Rec* 2008;291**:** 1173-1185.

8. Van Rikxoort EM, Isgum I, Arzhaeva Y, et al. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. *Medical Image Analysis* 2010;14**:**39-49.

9. Aljabar P, Heckemann RA, Hammers A, et al. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage* 2009;46:726-738.

10. Rohlfing T, Russakoff DB, Maurer CR. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. Ieee *Transactions on Medical Imaging* 2004;23: 983-994.

11. Crum WR, Griffin LD, Hill DLG, et al. Zen and the art of medical image registration: correspondence, homology, and quality. *Neuroimage* 2003;20:1425-1437.

12. Langerak TR, van der Heide UA, Kotte ANTJ, et al. Label Fusion in Atlas-Based Segmentation Using a Selective and Iterative Method for Performance Level Estimation (SIMPLE). *Ieee Transactions on Medical Imaging* 2010;29:2000-2008.

13. Artaechevarria X, Munoz-Barrutia A, Ortiz-de-Solorzano C. Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. Ieee *Transactions on Medical Imaging* 2009;28:1266-1277.

14. Coupe P, Manjon JV, Fonov V, et al. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage* 2011;54:940-954.

15. Ramus L, Commowick O, Malandain G. Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces. Medical Image Computing and Computer-Assisted Intervention. Jiang T, Navab, N, Pluim JPW, Viergever MA. Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces. Miccai 2010.

16. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys* 2014;41.

17. Langerak TR, Berendsen FF, Van der Heide UA, et al. Multiatlas-based segmentation with preregistration atlas selection. *Medical Physics* 2013;40.

18. Thiel W. Die Konservierung ganzer leichen in natürlichen farben. *Ann Anat*;174:185-195.

19. De Crop A, Bacher K, Van Hoof T, et al. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. *Radiology* 2012;262:298-304

20. Han X, Hibbard S, O'Connell N, et al. Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion. proc. *MICCAI Workshop Head & Neck Autosegmentation Challenge* 2010;297–304.

21. Taylor R. Interpretation of the correlation-coefficient - a basic review. *JDMS* 1990;6:35-39.

22. Kleinrensink GJ, Stoeckart R, Mulder PGH, et al. Upper limb tension tests as tools in the diagnosis of nerve and plexus lesions - Anatomical and biomechanical aspects. *Clin Biomech* 2000;15:9-14.

23. Coppieters MW, Stappaerts KH, Wouters LL, et al. Aberrant protective force generation during neural provocation testing and the effect of treatment in patients with neurogenic cervicobrachial pain. *JMPT* 2003;26:99-106.

24. Coppieters MW, Stappaerts KH, Staes FF, et al. Shoulder girdle elevation during neurodynamic testing: an assessable sign? *Manual Ther* 2001;6:88-96.

Supplementary material

To illustrate the differences between the similarity indices, their behavior with an increasing correctly delineated volume was investigated for 2 volumes with a 10 percent volume difference. The delineated volume was considered to be 10 percent larger than the gold standard volume, which is a realistic difference in volume with respect to the registered BPs. In figure 1 the results of this test are depicted.

First, the dice similarity coefficient (DSC) was calculated between the two segmentations (A, B). The DSC measures the spatial overlap between the gold standard A and the registered image B, and is defined as DSC(A,B)= 2(A∩B)/(A+B) where ∩ is the intersection volume. The DSC is situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

We also calculated the Jaccard similarity index (JI), as the ratio of intersection volume and the entire union volume of the delineations JI(A,B)= (A∩B)/(AUB) . The JI is also situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

Finally, the inclusion index (INI) was measured between the gold standard BP (A) and the registered BP (B). INI is the intersection volume of both, divided by the gold standard BP INI=(A∩B/A). INI is situated between 0 and 1 with 0 indicating no inclusion and 1 indicating total inclusion of A by B.

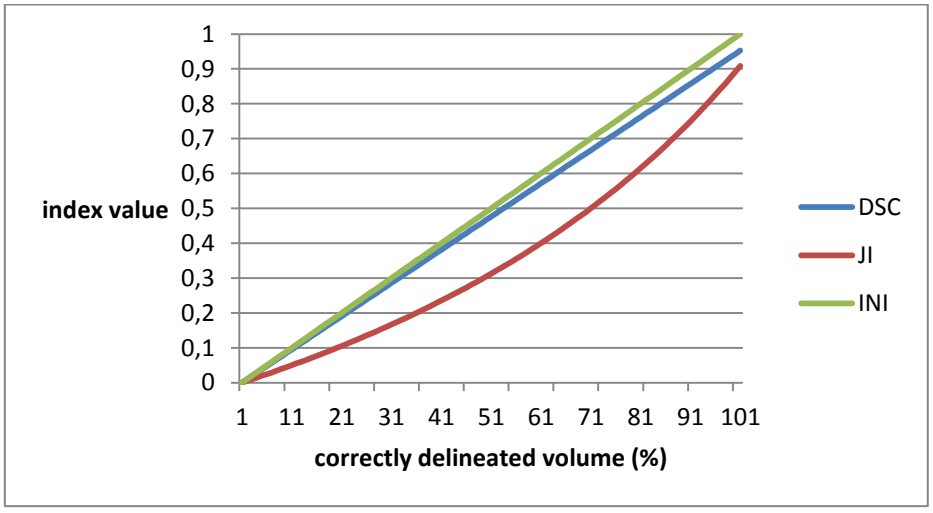


Fig. 1. Similarity indices as a function of correctly delineated volume. The behavior of the different similarity indices with increasing correctly delineated volume is illustrated for 2 volumes with a 10 percent volume difference. For illustrating the principle, the delineated volume (B) is randomly taken 10 percent larger than the gold standard volume (A) in this case. This is a realistic volume difference with respect to the registered BP's. DSC and INI both increase linearly with the increment of overlap between the 2 segmentations, whereas JI has a non-linear increment. The reason for this

difference is that with JI the penalty increases as the false positive delineation area increases. With DSC and INI in contrast, no penalty for false positive delineation area is given.

Manuscript 5

Optimal number of atlases and label fusion for automatic multi-atlas-based brachial plexus contouring in radiotherapy treatment planning

Joris Van de Velde, Johan Wouters, Tom Vercauteren, Werner De Gersem, Eric Achten, Wilfried De Neve, Tom Van Hoof

*JVdV and JW are shared first authors

Radiation Oncology. Published January 2016

A1, IF: 2.749

Abstract

Background: The present study aimed to define the optimal number of atlases for automatic multi-atlas-based brachial plexus (BP) segmentation and to compare Simultaneous Truth and Performance Level Estimation (STAPLE) label fusion with Patch label fusion using the ADMIRE® software. The accuracy of the autosegmentations was measured by comparing all of the generated autosegmentations with the anatomically validated gold standard segmentations that were developed using cadavers.

Materials and methods: Twelve cadaver computed tomography (CT) atlases were used for automatic multi-atlas-based segmentation. To determine the optimal number of atlases, one atlas was selected as a patient and the 11 remaining atlases were registered onto this patient using a deformable image registration algorithm. Next, label fusion was performed by using every possible combination of 2 to 11 atlases, once using STAPLE and once using Patch. This procedure was repeated for every atlas as a patient.

The similarity of the generated automatic BP segmentations and the gold standard segmentation was measured by calculating the average Dice similarity (DSC), Jaccard (JI) and True positive rate (TPR) for each number of atlases. These similarity indices were compared for the different number of atlases using an equivalence trial and for the 2 label fusion groups using an independent sample-t test.

Results: DSC's and JI's were highest when using 9 atlases with both STAPLE (average DSC= 0,532; JI= 0,369) and Patch (average DSC= 0,530; JI= 0,370). When comparing both label fusion algorithms using 9 atlases for both, DSC and JI values were not significantly different. However, significantly higher TPR values were achieved in favour of STAPLE ($p< 0,001$). When fewer than 4 atlases were used, STAPLE produced significantly lower DSC, JI and TPR values than did Patch ($p=0,0048$).

Conclusions: Using 9 atlases with STAPLE label fusion resulted in the most accurate BP autosegmentations (average DSC= 0,532; JI= 0,369 and TPR= 0,760). Only when using fewer than 4 atlases did the Patch label fusion results in a significantly more accurate autosegmentation than STAPLE.

Keywords: automatic, multi-atlas-based, brachial plexus , segmentation, label fusion, cadavers

Background

Manual brachial plexus (BP) segmentation on planning computed tomography (CT) for radiation therapy treatment planning is a time-consuming and unreliable process [1]. An effective automatic BP contouring method could relieve clinicians of this tedious task and would result in a higher inter- and intra-observer reliability and accuracy of the contouring process [2]. This issue is of growing importance following the introduction of both function-sparing and adaptive intensity modulated radiation therapy (IMRT), where the number and frequency of delineation of organs at risk (OAR's) are increased [3, 4].

The BP is one of the OAR's in IMRT for head-and-neck, lung and breast cancer patients. However, in clinical practice, the BP is often not delineated during treatment planning, and when the delineation is included, the accuracy tends to be low [1]. Incorrect or absent delineations of OAR's in IMRT treatment planning however, have proven to be a main source of uncertainty in historical dose-volume effect data, which leads to the reduced performance of predictive models [5]. Moreover, when the BP radiation dose is not controlled during IMRT treatment, the possibility exists that the dose to this organ exceeds the BP tolerance dose of 66 Gy [6, 7, 8], which can potentially cause radiation-induced brachial plexopathy (RIBP). RIBP was thought to be uncommon for head-and-neck cancer patients, but recent clinical investigations have suggested that it remains underreported [9, 10].

To develop accurate automatic multi-atlas-based BP segmentations, multiple parameters must be controlled.

The first parameter is the optimal image registration and label fusion algorithm that are used. In multi-atlas-based autosegmentation strategies, several available presegmented images –called atlases– are first registered separately to the patient using deformable image registration. During the deformable image registration process a deformation vector field (DVF), describing the non-linear transformation from a presegmented image dataset to a patient image dataset, is created. Based on the computed DVF a set of delineations on the presegmented image data set are deformed on the patient image data set. The series of deformed delineations on the patient image data set are combined by the label fusion algorithm to obtain a unique and final consensus segmentation. Multiple image registration and label fusion algorithms for various organs have been compared in the literature [11, 12, 13, 14]. However, only one publication [15] investigated BP autosegmentation. In that study, the authors concluded that multi-atlas autosegmentation can be effectively used to delineate BP on CT. However, these conclusions may be unreliable because the autosegmentation itself and also the subsequent validation procedure were based on BP gold standards that were not validated [1]. Moreover, Yang et al. [15] used the Simultaneous Truth and Performance Level

Estimation (STAPLE) algorithm to generate automatic BP contours. A described weakness of the STAPLE label fusion method is that it ignores the image data and uses only the segmentations when computing the label fusion [16]. In an attempt to counter this weakness, an additional intensity weighted label fusion method called 'Patch' fusion [17] was recently implemented in the ADMIRE® software. The ADMIRE® white paper reported that this Patch fusion outperforms the STAPLE algorithm for some anatomical structures with a very stable anatomical topography [18]. No publication has been found, however, in which the use of this label fusion algorithm was investigated for automatic multi-atlas-based BP segmentation.

A second parameter that must be controlled to obtain the maximum accuracy is the number of atlases that has to be used for multi-atlas-based BP autosegmentation. In different publications, multi-atlas-based automatic segmentation methods have proven to be more effective than single-atlas-based methods [19, 20] but the specific number of atlases to use was investigated in only a few publications [21, 14]. None of these studies provided a specific number for optimal automatic BP contouring.

The purpose of this study was to define the optimal number of atlases to use for automatic multi-atlas-based BP contouring and to compare the STAPLE algorithm with Patch label fusion using the ADMIRE® software. This was measured by comparing all of the generated automatic BP segmentations with high-quality, anatomically validated, gold standard atlases that were developed using cadavers.

Materials and methods

To develop gold standard atlases for BP contouring, 12 cadavers (age and gender randomized) were used. The cadavers were embalmed according to Thiel because of their optimal image quality and movement capacities [22, 23]. The latter allowed for the required standardization of the scan position. Magnetic resonance imaging (MRI) of the head-and-neck region was performed to generate high-quality BP delineations that were anatomically validated by dissection. These anatomically validated, MRI-based, BP delineations were then rigidly fused to the corresponding CT to obtain BP gold standard delineations that were applicable to the radiation therapy planning system. A detailed description was provided by Van de Velde et al. [24]. This study was approved by the ethics committee of University Hospital Ghent (reference number: B67020142069), and was in compliance with the Helsinki Declaration.

For image registration and label fusion, the ADMIRE® software 1.10.02 (Elekta AB, Stockholm, Sweden) was used. ADMIRE® performs the segmentation of a novel subject image (here called

'patient') by using multiple pre-segmented images, which are also known as 'atlases'. The 'General' algorithm in ADMIRE® is used for the initial deformable image registration [18]. This image registration framework consists of three major steps: a linear registration and two non-linear registration steps. With each step the number of degrees of freedom increases, and is used to provide initialization for the next step.

For label fusion, 2 different algorithms in ADMIRE® are compared: the STAPLE label fusion [16] and Patch label fusion [17]. The STAPLE algorithm works with a statistical framework that simultaneously estimates the underlying 'truth' segmentation and the accuracy of each individual atlas [18]. It ignores the image data and uses only the segmentations when computing the label fusion. In contrast, the Patch algorithm considers the accuracy of the initial image registration by comparing the intensity similarity between the atlas and the patient after being aligned, to get better label fusion results. This process, is called 'intensity weighting'.

Procedure

The present study aimed to determine the optimal number of atlases and to compare the STAPLE with the Patch label fusion algorithm for multi-atlas-based BP contouring in ADMIRE® software.

For this purpose, a leave-one-out strategy was followed. One of the 12 available cadaver CT-datasets was selected as a patient and the remaining CT-datasets, which contained the anatomically validated BP segmentation, served as atlases. All of the atlases were first registered separately onto the patient using the 'General' registration algorithm in ADMIRE®. Next, the label fusion was performed, with both STAPLE and Patch, first using every possible combination of 2 atlases. Subsequently, label fusion was repeated with a gradually increasing number of atlases, until every possible combination of 11 atlases was reached. This process was reiterated for every atlas as a patient. It resulted in 24432 combinations over the different number of atlases. A Power analysis was executed (power π= 80) to calculate the minimum sample size required for a 90% confidence interval.

Next, for every generated 'label fused' autosegmentation, 3 similarity indices with the gold standard contour were calculated to quantify the accuracy (Table 1):

First, Dice similarity coefficient (DSC) was calculated between these 2 segmentations. The DSC measures the spatial overlap between the gold standard A and the registered image B, and is defined as $DSC(A,B) = 2(A \cap B)/(A+B)$ where $\cap$ is the intersection volume. The DSC is situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

We also calculated the Jaccard index (JI) as the ratio of the intersection volume and the entire union volume of the delineations: $JI(A,B) = (A \cap B)/(AUB)$. The JI is also situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

At last, True positive rate (TPR) was measured between the gold standard BP (A) and the registered BP (B). TPR is the intersection volume of these, divided by the gold standard BP: TPR = (A∩B/A). TPR is situated between 0 and 1 with 0 indicating no inclusion and 1 indicating the total inclusion of A by B.

Finally, for each number of atlases, average DSC, JI and TPR were calculated over the different combinations.

To determine the clinically relevant optimal number of atlases, an equivalence trial was conducted [25, 26]. An equivalence trial is used to demonstrate similarity between compared groups. It uses a confidence interval in which equivalence is claimed when the confidence interval of the difference in outcome between compared groups is within a predetermined equivalence margin. This equivalence margin represents a clinically acceptable range of differences. For this study, an equivalence margin of 10 percent was predetermined.

Only DSC and JI were appropriate as a reference for the equivalence trial, because in those indices, the most accurate segmentation will be associated with the highest index values, since both indices consider a penalty for false positive delineation area. The TPR from its side was not adequate for the equivalence trial because the highest TPR value does not necessary imply the most accurate segmentation [27], since a false positive delineation area is not penalized in this index.

DSC was chosen for equivalence trial over JI because the DSC has a linear course with an increasing correctly delineated volume and JI has not. Thus, a 10 percent (= equivalence margin) increase or decrement of DSC always correlates with the same amount of increase or decrement of the correctly delineated volume [27]. Using JI conversely, the amount of correctly delineated volume associated with an increase or decrease of 10 percent JI value, will vary depending on the starting value of the JI, because this index has a non-linear course. For example, an increase in JI value from 0.8 to 0.9 will result in a larger increase in percentage of correctly delineated volume than an increase from 0.2 to 0.3 [27].

Starting from the number of atlases with the maximal DSC values (reference group), the number of atlases was first gradually increased by one. If, by increasing the number of atlases each time starting from the reference group, the decrease of DSC (90% CI) felt within the equivalence margin of 10 percent, the groups were considered to be equivalent. This procedure was performed for the two label fusion groups separately [26]. Only in case of equivalent DSC values combined with significantly higher TPR values, the autosegmentation result was considered to be more accurate, because in this case the equivalence of the DSC values indicates that the increase of the false positive delineation area, which is not penalized by TPR, was kept within bounds.

Next, the number of atlases was gradually decreased by one, starting from the reference group. If, by decreasing the number of atlases each time starting from the reference group, the

decrement of the DSC values fell within the equivalence margin, the calculation time could be reduced by using a lower number of atlases without clinically relevant loss in accuracy.

Thereafter, the difference between STAPLE and Patch label fusion was determined using an independent sample t-test. Therefore, in the 2 label fusion groups, the similarity indices for their respective clinically relevant optimal number of atlases were compared.

Results

The power analysis (π= 80) resulted in a sample size of 150 combinations per number of atlases needed for a 90% confidence interval. For each number of atlases, the average DSC, JI and TPR, their standard deviations and their possible combinations (samples) are shown in table 1 and figures 2 and 3 for both groups.

Table 1. Average Dice similarity coefficient, Jaccard index and True positive rate per number of atlases.

		STAPLE			Patch		
Number of atlases	**Samples**	**DSC (SD)**	**JI (SD)**	**TPR (SD)**	**DSC (SD)**	**JI (SD)**	**TPR (SD)**
2	660	0,247 (0,179)	0,154 (0,131)	0,188 (0,158)	0,400 (0,157)	0,262 (0,124)	0,416 (0,178)
3	660	0,397 (0,184)	0,265 (0,151)	0,373 (0,187)	0,454 (0,157)	0,307 (0,136)	0,439 (0,163)
4	3960	0,472 (0,171)	0,325 (0,147)	0,473 (0,184)	0,477 (0,161)	0,328 (0,141)	0,445 (0,165)
5	5544	0,482 (0,153)	0,331 (0,132)	0,534 (0,166)	0,465 (0,149)	0,316 (0,128)	0,435 (0,150)
6	5544	0,519 (0,138)	0,362 (0,128)	0,616 (0,155)	0,501 (0,146)	0,347 (0,133)	0,465 (0,150)
7	3960	0,514 (0,129)	0,356 (0,117)	0,658 (0,147)	0,492 (0,144)	0,339 (0,131)	0,446 (0,142)
8	1980	0,501 (0,120)	0,343 (0,106)	0,686 (0,143)	0,501 (0,140)	0,346 (0,127)	0,466 (0,140)

9	660	**0,532***	**0,369***	0,726	**0,530***	**0,370***	0,466
		(0,102)	**(0,940)**	(0,127)	**(0,117)**	**(0,112)**	(0,125)
10	132	0,510	0,349	0,742	0,524	0,365	0,468
		(0,100)	(0,900)	(0,127)	(0,124)	(0,116)	(0,121)
11	12	0,506	0,344	**0,760***	0,530	0,370	**0,471***
		(0,940)	(0,840)	**(0,126)**	(0,122)	(0,115)	**(0,115)**

Abbreviations: DSC, Dice similarity coefficient; JI, Jaccard index; INI, Inclusion index

**Highest index values*

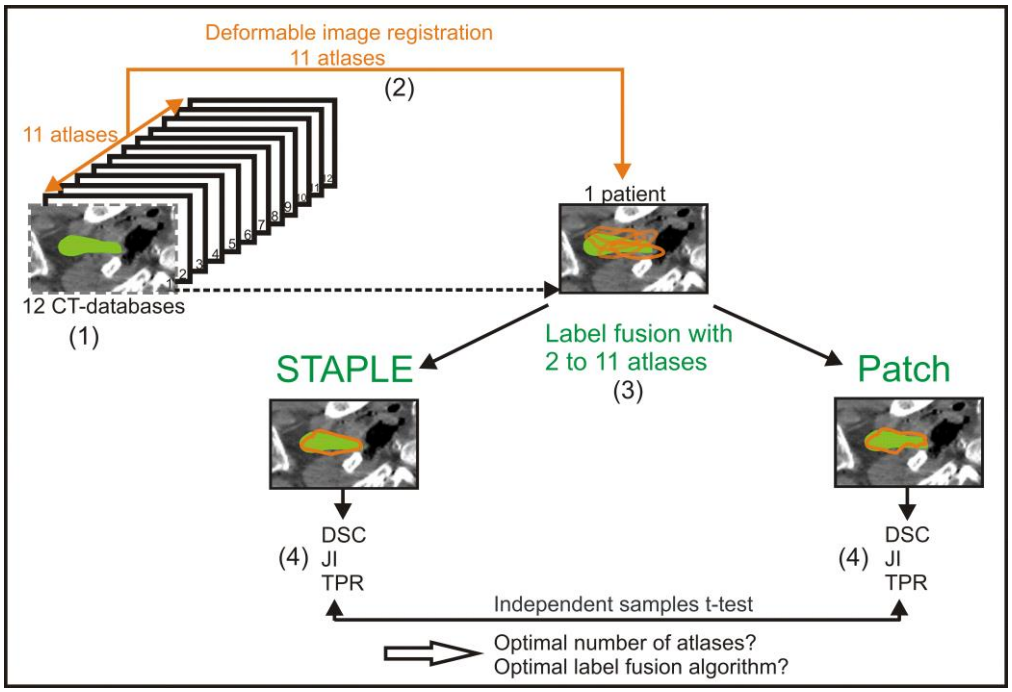


Fig. 1. Schematic illustration of the procedure for determining the optimal number of atlases and optimal label fusion. (1) Twelve cadaver CT datasets were included, and one atlas was selected as a patient. (2) The 11 remaining atlases were used for deformable image registration on the patient. (3) Label fusion was performed with 2 up to 11 atlases, once using STAPLE and once using Patch. (4) For each number of atlases, the average Dice similarity coefficient (DSC), Jaccard index (JI) and True positive rate (TPR) were calculated for the generated contour (orange) with the gold standard contour (green). This procedure was repeated for every atlas as a patient.

The highest average DSC and JI values were found when using 9 atlases for both STAPLE and Patch fusion (Table 1, Fig. 1). When more than 9 atlases were used, only TPR for STAPLE label fusion continued to increase (Fig. 2).

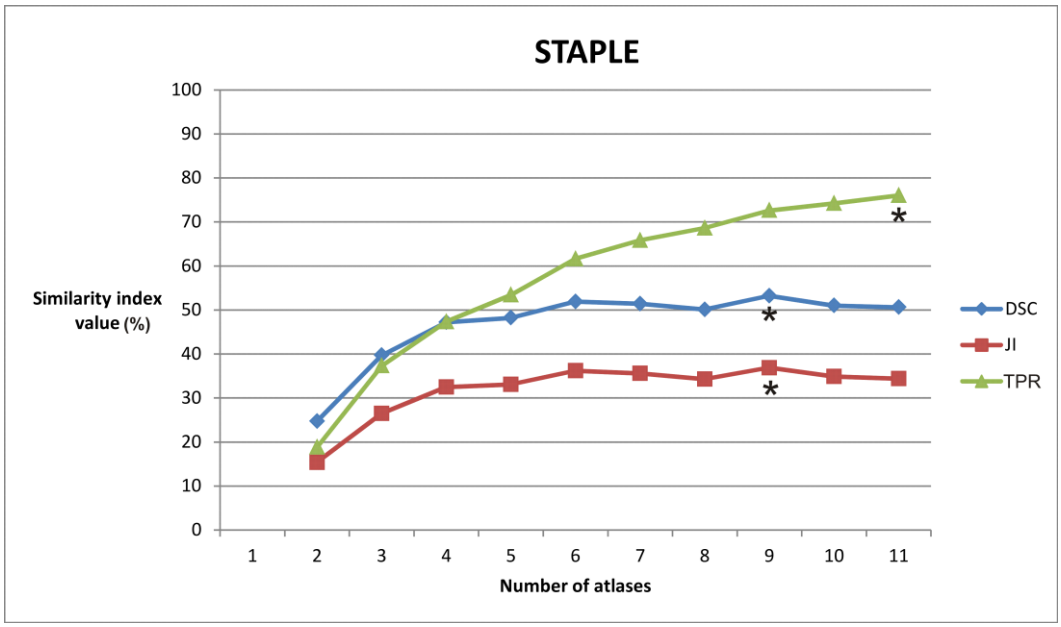


Fig. 2. Behaviour of the average similarity indices with an increasing number of atlases fused using the STAPLE algorithm. * indicates the highest similarity index values. DSC, Dice similarity coefficient; JI, Jaccard index; TPR, true positive rate.

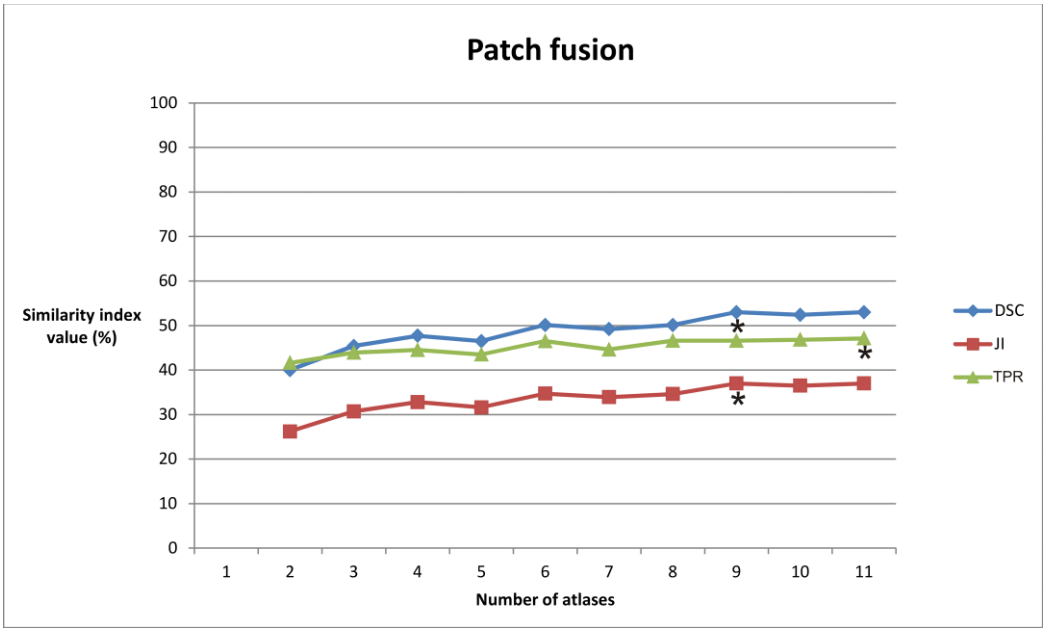


Fig. 3. Behaviour of the average similarity indices with an increasing number of atlases fused using the Patch algorithm. * indicates the highest similarity index values. DSC, Dice similarity coefficient; JI, Jaccard index; TPR, true positive rate.

By increasing the number of atlases for STAPLE label fusion from 9 to 10 atlases, the decrease in DSC values still fell within the predisposed equivalence margin of 10 percent (Fig. 4) but no significantly higher TPR values were achieved. When using 11 atlases, DSC were no longer equivalent to the results obtained when using 9 atlases. However, the number of possible combinations for a power of 90 (sample size of 150) was not sufficient with 10 and 11 atlases, so no definitive conclusions can be drawn concerning these number of atlases.

By decreasing the number of atlases from 9 to 6 atlases the decrease in DSC values still fell within the predisposed equivalence margin (Fig. 4). The average calculation time was reduced from 19 min to 17 min. When using fewer than 6 atlases, DSC values were no longer equivalent to the results obtained when using 9 atlases.

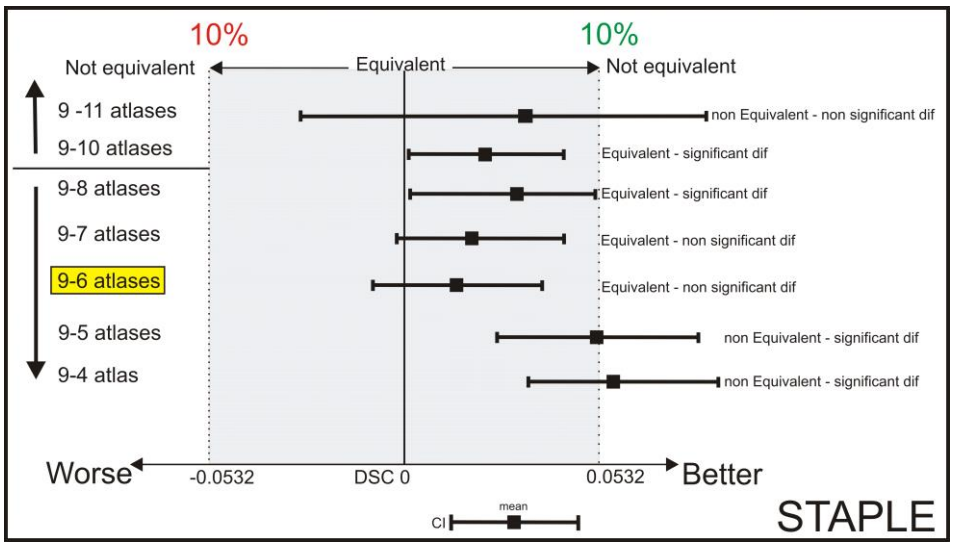


Fig. 4. Interpretation of the equivalence using 9 atlases in multi-atlas brachial plexus autosegmentation compared to using more (10-11) and fewer atlases (8-7-6-5-4) with STAPLE label fusion. Using 6 atlases is equivalent to using 9 atlases (yellow marked). The shaded area covers the equivalence range of 10%. □ = observed point estimate of the outcome difference in each number of atlases, corresponding error bar= two-sided 90% confidence interval (caps at each end = lower and upper bar bounds of confidence interval). DSC, Dice similarity coefficient; CI, confidence interval.

By increasing the number of atlases for Patch label fusion from 9 to 10 atlases, the decrease in the DSC values did fall within the predisposed equivalence margin of 10 percent (Fig. 5) but no significantly higher TPR values were achieved. When the number of atlases was decreased until 8 or lower, the decrease in DSC was not within the equivalence margin (Fig. 5). Also here, the number of possible combinations for a power of 90 (sample size of 150) was not sufficient with 10 and 11 atlases.

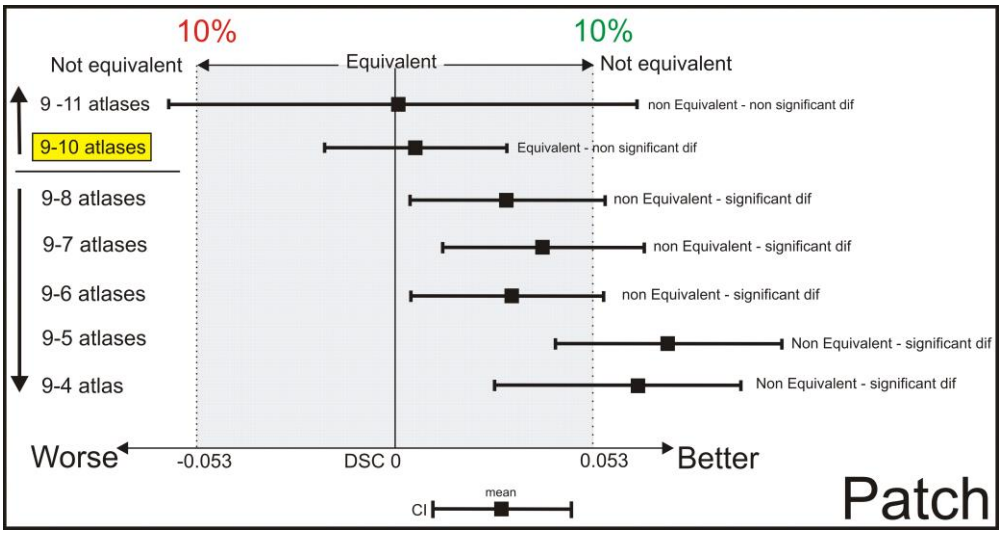


Fig. 5. Interpretation of the equivalence using 9 atlases in multi-atlas brachial plexus autosegmentation compared with using more (10-11) and fewer (8-7-6-5-4) atlases with Patch label fusion. Using 10 atlases is equivalent to using 9 atlases (yellow marked). The shaded area covers the equivalence range of 10%. □ = observed point estimate of the outcome difference in each number of atlases, corresponding error bar= two-sided 90% confidence interval (caps at each end = lower and upper bar bounds of confidence interval). DSC, Dice similarity coefficient; CI, confidence interval.

When comparing the label fusion algorithms using the optimal number of atlases for both (9 for both STAPLE and Patch) DSC and JI values were not significantly different. However, significantly higher TPR values were achieved in favour of STAPLE ($p< 0,001$).

Only when fewer than 4 atlases were used, STAPLE resulted in significantly lower DSC, JI and TPR values than Patch ($p=0,0048$).

Discussion

The purpose of this study was to determine the optimal number of atlases for automatic multi-atlas-based BP contouring and to compare STAPLE and PATCH label fusion algorithms using the ADMIRE® software.

For STAPLE, the average DSC and JI values were maximal when using 9 atlases (Table 1). When the number of atlases was increased to 10, the DSC values remained equivalent to those obtained when using 9 atlases, but the TPR values were not significantly higher.

The most accurate autosegmentation results are achieved when the JI and DSC values reach their maximum value. The number of atlases where these maximum values are reached are identical for both similarity indices. From this point on, it is only possible to achieve higher TPR values by adding more atlases. However, when the TPR values increase and the DSC and JI values decrease, the increase in the true-positive delineation area is associated with a proportionally larger increase in the false positive delineation area. This may occur because TPR only measures the increase in the true-

positive delineation area and does not penalize an increase of the false positive delineation area. DSC and JI, in contrast, do penalize an increase in the false-positive delineation area.

Because the decrease in DSC from 9 to 10 atlases fell within the equivalence margin, and because the TPR values were not significantly higher, autosegmentations obtained with 10 atlases could not improve the accuracy compared to those obtained using 9 atlases for STAPLE. Consequently, the optimal number of atlases is 9 for STAPLE label fusion.

In the case of limited computer calculation power, 6 atlases could be used for STAPLE without a clinically relevant loss of accuracy and an average time saving of 2 min.

For Patch, the DSC and JI values were also maximal at 9 atlases. By increasing the number of atlases to 10, no significant increase in TPR values was achieved either, which indicated that Patch fusion with 9 atlases also resulted in the most accurate autosegmentations. To reduce the calculation time, the number of atlases cannot be decreased without a clinically relevant loss of autosegmentation accuracy (Fig. 5).

Comparing both label fusion algorithms (STAPLE and Patch) using their respective optimal number of atlases (9 for both), DSC and JI values were not significantly different. However, significantly higher TPR values were found in favour of STAPLE ($p < 0,001$). Therefore, we recommend using STAPLE label fusion with 9 atlases to obtain the most accurate autosegmentations results.

Conversely, when fewer than 4 atlases were used, STAPLE provided significantly less accurate results than did Patch ($p=0.004862$). So, Patch label fusion is preferable over STAPLE when only less than 4 atlases are available.

The current study is the first to investigate the optimal number of atlases for BP autosegmentation. The optimal number of atlases for some other organs was already studied: for the nucleus caudatus Aljabar et al. (2009) [21] concluded that using 8 atlases is optimal; for the hippocampus, the highest accuracy is reached with a selection of 25 atlases. Pirozzi et al. (2012) concluded that for the bladder and the femur, the optimal number of atlases was 5, and that the optimal number for the prostate and rectum was 4 [14]. Remarkable is that in the first study [21], the number of atlases for autosegmentation of anatomically stable brain structures is higher than in the second study [14], in which anatomically variable organs were autosegmented. The opposite could be expected. The varying results of these studies only show that the optimal number of atlases is very organ-dependent and especially algorithm-dependent. So, for more general conclusions concerning the optimal number of atlases for BP autosegmentation, other algorithms also have to be investigated.

Few studies were found that compared ADMIRE® software to other autosegmentation software. Simmat et al. [11] found higher flexibility and robustness in the algorithm used in the ADMIRE® software compared with the algorithms in Iplan®[12] for the bladder, prostate and rectum. La

Macchia et al. [13] found the best label fusion results using STAPLE in ADMIRE® compared with the algorithms in VelocityA® and MIM 5®, for the head-and-neck region. BP autosegmentation was not included in both studies. For general conclusions concerning the best autosegmentation software for BP autosegmentation, different autosegmentation software need to be compared in further studies.

Another limitation of the current study is that only 12 atlases were available. Hence, for the combinations with 10 and 11 atlases, the sample size was not big enough to draw definitive conclusions. To increase the statistical power and to draw definitive conclusions for the highest number of atlases, more atlases need to be included in the study. The more atlases included in the study, the more accurate the autosegmentation results will be as well, because the probability of selecting atlases that are more similar to the patient's morphotype will increase.

The dosimetric implications of optimization of label fusion and the number of atlases on radiation therapy treatment planning were not included in the investigation. Additional studies are in process to study the dosimetric impact and measure the potential benefit for patients undergoing radiation therapy treatment. Future perspectives include further increasing the accuracy of the automatic BP segmentations to a clinically acceptable level, by combining the optimal number of atlases and label fusion with an effective atlas selection strategy and including higher number of anatomically validated atlases to study the effect of using more than 11 atlases.

Conclusion

STAPLE is preferable to Patch label fusion for multi-atlas-based BP autosegmentation. Only when fewer than 4 atlases are available, it's preferable to choose Patch above STAPLE.

Using 9 atlases with STAPLE resulted in the most accurate BP autosegmentations. With a limited computer calculation power, the number of atlases could be decreased until 6 without a clinically relevant loss of accuracy.

List of abbreviations

BP, Brachial plexus; STAPLE, Simultaneous truth and performance level estimation; CT, Computed tomography; DSC, Dice similarity coefficient; JI, Jaccard index; INI; Inclusion index; IMRT, Intensity modulated radiation therapy; OAR, Organs at risk; MRI, Magnetic resonance imaging.

Competing interests

The authors declare that they have no competing interests.

Author contributions

JVDV and JW contributed to the conception and design, acquisition of data and analysis, interpretation of data and drafting of the manuscript. TV was involved in the acquisition of the data, software development and drafting of the manuscript. WDS was involved in the acquisition of the data, software development and drafting of the manuscript. EA contributed to the conception and design and drafting of the manuscript. WDN contributed to the conception and design and drafting of the manuscript. TVH coordinated the investigation, was involved in the acquisition of data and analysis, interpretation of data and drafting of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors thank Elekta AB, Stockholm, Sweden.

This research was supported by the Foundation against Cancer, grant 2012-200.

References

1. Van de Velde J, Vercauteren T, De Gersem W, Wouters J, Vandecasteele K, Vuye P, et al. Reliability and accuracy assessment of radiation therapy oncology group-endorsed guidelines for brachial plexus contouring. Strahlenther Onkol. 2014;190:628-635.
2. Stapleford LJ, Lawson JD, Perkins C, Edelman S, Davis L, McDonald MW, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. Int J Radiat Oncol Biol Phys. 2010;77:959-966.
3. Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. Radiat Onco. 2014;9(1):173.
4. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. Med Phys. 2014, 41.
5. Brouwer CL, Steenbakkers RJ, Van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D variation in delineation of head and neck organs at risk. Radiat Oncol. 2012;7:32.
6. Truong M, Romesser PB, Qureshi MM, Orlina L, Willins JD. Dose Volume Histogram Analysis of Routine Brachial Plexus Contouring for Head and Neck Intensity Modulated Radiation Therapy. Int J Radiat Oncol Biol Phys. 2011;81:S532-S532.

7. Hall WH, Guiou M, Lee NY, Dublin A, Narayan S, Vijayakumar S, et al. Development and validation of a standardized method for contouring the brachial plexus: preliminary dosimetric analysis among patients treated with imrt for head-and-neck cancer. Int J Radiat Oncol Biol Phys. 2008;72:1362-1367.
8. Kong FM, Ritter T, Quint DJ, Senan S, Gaspar LE, Komaki RU, et al. consideration of dose limits for organs at risk of thoracic radiotherapy: atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. Int J Radiat Oncol Biol Phys. 2011;81:1442-1457.
9. Chen AM, Hall WH, Li J, Beckett L, Farwell DG, Lau DH,et al. Brachial Plexus-Associated Neuropathy After High-Dose Radiation Therapy for Head-and-Neck Cancer. Int J Radiat Oncol Biol Phys. 2012;84:165-169.
10. Mutter RW, Lok BH, Dutta PR, et al. Constraining the brachial plexus does not compromise regional control in oropharyngeal carcinoma. Radiat Oncol. 2013;8:173.
11. Simmat I, Georg P, Georg D, Birkfellner W, Goldner G, Stock M. Assessment of accuracy and efficiency of atlas-based autosegmentation for prostate radiotherapy in a variety of clinical conditions. Strahlenther Onkol. 2012;188:807-813.
12. Noblet V, Heinrich C, Heitz F, Armspach JP. Retrospective evaluation of a topology preserving non-rigid registration method. Med Image Anal. 2006;10:366-384.
13. La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. Radiat Oncol. 2012;7:160.
14. Pirozzi, Horvat M, Piper J, et al. Atlas-based Segmentation: Evaluation of a Multi-Atlas Approach for Lung Cancer. Med Phys. 2012;39:3677.
15. Yang J, Amini A, Williamson R, et al. Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy. PRO. 2013;3:139-147.
16. Warfield SK, Zou KH, Wells WM. Validation of image segmentation by estimating rater bias and variance. Philos Trans A Math Phys Eng Sci. 2008;366:2361-2375.
17. Coupe P, Manjon JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage 2011;54:940-954.
18. Han X. Elekta AB (2012). A Locally Adaptive, Intensity-Based Label Fusion Method for Multi-Atlas Auto-Segmentation. Maryland Heights, MO.
19. Teguh DN, Levendag PC, Voet PWJ, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue

(swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys. 2011;81:950-957.

20. Anders LC, Stieler F, Siebenlist K, Schafer J, Lohr F, Wenz F. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. Radiother Oncol. 2012;102:68-73.
21. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage. 2009;46:726-738.
22. Thiel W. Die Konservierung ganzer leichen in natürlichen farben. Ann Anat. 1992;174:185-195.
23. De Crop A, Bacher K, Van Hoof T, Smeets PV, Smet BS, Vergauwen M, et al. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. Radiology. 2012;262:298-304
24. Van de Velde J, Audenaert E, Speleers B, Vercauteren T, Mulliez T, Vandemaele P, et al. An Anatomically Validated Brachial Plexus Contouring Method for Intensity Modulated Radiation Therapy Planning. Int J Radiat Oncol Biol Phys. 2013;87:802-808.
25. Veronesi U, Orecchia R, Maisonneuve P, Viale G, Rotmensz N, Sangalli C, et al. Intraoperative radiotherapy versus external radiotherapy for early breast cancer (ELIOT): a randomised controlled equivalence trial. Lancet Oncology. 2013;14:1269-1277.
26. Ahn S, Park SH, Lee KH. How to Demonstrate Similarity by Using Noninferiority and Equivalence Statistical Testing in Radiology Research. Radiology. 2013;267:328-338.
27. Van de Velde J, Wouters J, Vercauteren T, De Gersem W, Duprez F, De Neve W, et al. Morphometric Atlas Selection for Automatic Brachial Plexus Segmentation. Int J Radiat Oncol Biol Phys. 2015; doi:10.1016/j.ijrobp.2015.02.045.

Manuscript 6

The effect of morphometric atlas selection on multi-atlas based automatic brachial plexus segmentation

Joris Van de Velde, Johan Wouters, Tom Vercauteren, Werner De Gersem, Eric Achten, Wilfried De Neve, Tom Van Hoof

*JVdV and JW are shared first authors

Radiation Oncology. Published December 2015.

A1, IF: 2.749

Abstract

Purpose: The present study aimed to measure the effect of a morphometric atlas selection strategy on the accuracy of multi-atlas-based BP autosegmentation using the commercially available software package ADMIRE® and to determine the optimal number of selected atlases to use. Autosegmentation accuracy was measured by comparing all generated automatic BP segmentations with anatomically validated gold standard segmentations that were developed using cadavers.

Materials and methods: Twelve cadaver computed tomography (CT) atlases were included in the study. One atlas was selected as a patient in ADMIRE®, and multi-atlas-based BP autosegmentation was first performed with a group of morphometrically preselected atlases. In this group, the atlases were selected on the basis of similarity in the shoulder protraction position with the patient. The number of selected atlases used started at 2 and increased up to 8. Subsequently, a group of randomly chosen, non-selected atlases were taken. In this second group, every possible combination of 2 to 8 random atlases was used for multi-atlas-based BP autosegmentation. For both groups, the average Dice similarity coefficient (DSC), Jaccard index (JI) and Inclusion index (INI) were calculated, measuring the similarity of the generated automatic BP segmentations and the gold standard segmentation. Similarity indices of both groups were compared using an independent sample t-test, and the optimal number of selected atlases was investigated using an equivalence trial.

Results: For each number of atlases, average similarity indices of the morphometrically selected atlas group were significantly higher than the random group ($p < 0,05$). In this study, the highest similarity indices were achieved using multi-atlas autosegmentation with 6 selected atlases (average DSC= 0,598; average JI= 0,434; average INI= 0,733).

Conclusions: Morphometric atlas selection on the basis of the protraction position of the patient significantly improves multi-atlas-based BP autosegmentation accuracy. In this study, the optimal number of selected atlases used was 6, but for definitive conclusions about the optimal number of atlases and to improve the autosegmentation accuracy for clinical use, more atlases need to be included.

Key words

morphometric atlas selection, brachial plexus, autosegmentation, multi-atlas-based, cadavers

Background

Recent technological and procedural advances in radiotherapy, such image guided radiotherapy and adaptive replanning, require numerous image acquisitions and a subsequent delineation of target structures and organs at risk (OAR). Repetitive delineation of OARs is very tedious and time-consuming, and it can only be facilitated by automatic segmentation software [1]. The need for automatic segmentation software certainly applies to OARs, which are especially difficult to segment due to their poor visibility on planning-CT. One of these invisible OARs in patients with lung, breast and head-and-neck cancer is the brachial plexus (BP).

Accurate automatic multi-atlas-based BP segmentations for radiotherapy treatment planning are hard to achieve [2]. Several parameters need to be controlled in order to obtain a clinically reliable automatic BP segmentation. Optimal deformable image registration and label fusion algorithms need to be chosen, the optimal number of atlases need to be determined, and the most patient-similar atlases need to be selected for registration. In earlier publications [3], the best label fusion algorithm and optimal number of atlases for automatic BP segmentation without atlas selection were already determined in ADMIRE® software. However, the effect of implementation of an atlas selection strategy on multi-atlas-based BP autosegmentation accuracy has not yet been investigated.

Several organ-aspecific atlas preselection strategies to select the most suitable atlases for multi-atlas-based automatic contouring have been previously published. Most of these strategies are based on similarities between the atlas and the patient image [4, 5]. Also meta-information, such as body mass index, age, pathology, clinical history, gender and handedness can be used as selection criteria [6]. The weakness of image similarity selection methods, which use largely irrelevant areas of anatomical information, and the weakness of meta-information selection methods, which cannot address anatomical variability, could possibly be countered by using only organ-specific stable anatomical information in the registration process.

An organ-specific atlas preselection strategy based on morphometric parameters was developed for the BP by Van de Velde et al. [3]. The authors measured several morphometric parameters related to the BP on atlas- and patient-Computed Tomography (CT) and investigated which parameters significantly improved the autosegmentation result. A clinically relevant effect was found using atlas selection based on the protraction difference between the atlas and the patient on single-atlas based automatic BP segmentation. The effect of morphometric atlas selection on the accuracy of multi-atlas-based BP autosegmentation has not yet been investigated, nor the optimal number of selected atlases to use.

The present study aims to measure the effect of a morphometric atlas selection strategy implementation on the accuracy of multi-atlas-based BP autosegmentation and to determine the optimal number of selected atlases to use. Segmentation accuracy was measured by comparing the generated automatic BP segmentations with high quality anatomically validated gold standard atlases that were developed using cadavers [7].

Material and methods

To develop gold standard atlases for BP contouring, 12 cadavers (age and gender randomized) were used. The cadavers were embalmed according to Thiel because of their optimal image quality and movement capacities [8, 9]. The latter allowed for the required standardization of the scan position. Magnetic resonance imaging (MRI) of the head-and-neck region was performed to generate high-quality BP delineations that were anatomically validated by dissection. These anatomically validated, MRI-based BP delineations were then rigidly fused to the corresponding CT to obtain BP gold standard delineations that were applicable to the radiation therapy planning system. A detailed description of this method was provided by Van de Velde et al. [7]. The current study was approved by the ethics committee of the University Hospital Ghent (reference number: B67020142069) and was in compliance with the Helsinki Declaration.

The commercially available software package ADMIRE® 1.10.02 (Elekta AB, Stockholm, Sweden) was used for multi-atlas-based autosegmentation.

In multi-atlas-based autosegmentation strategies, several available presegmented images –called atlases– are first separately registered to the patient using deformable image registration. During the deformable image registration process, a deformation vector field (DVF) describing the non-linear transformation from a presegmented image dataset to a patient image dataset is created. Based on the computed DVF, a set of delineations on the presegmented image data set are deformed on the patient image data set. The multiple deformed delineations on the patient image data set are combined by the label fusion algorithm to obtain a unique and final consensus segmentation.

In ADMIRE®, 2 label fusion algorithms can be applied: the Simultaneous Truth and Performance Level Estimation (STAPLE) [10] and Patch [11] label fusion. STAPLE was proven to be more effective than Patch label fusion for multi-atlas-based BP autosegmentation [3]. This label fusion algorithm was originally designed for the validation of image segmentations. It considers a collection of

segmentations and computes a probabilistic estimate of the true segmentation, as well as a measure of the performance level represented by each segmentation [10, 12].

Procedure

In the first step, in order to measure the effect of a morphometric atlas selection strategy implementation in multi-atlas-based BP segmentation, the protraction distances were measured for all atlases [3]. The protraction distance is defined as the horizontal distance measured in the sagittal plane between a vertical line through the anterior tubercle of C5 and a vertical line through the infraglenoid tubercle (Fig. 1).

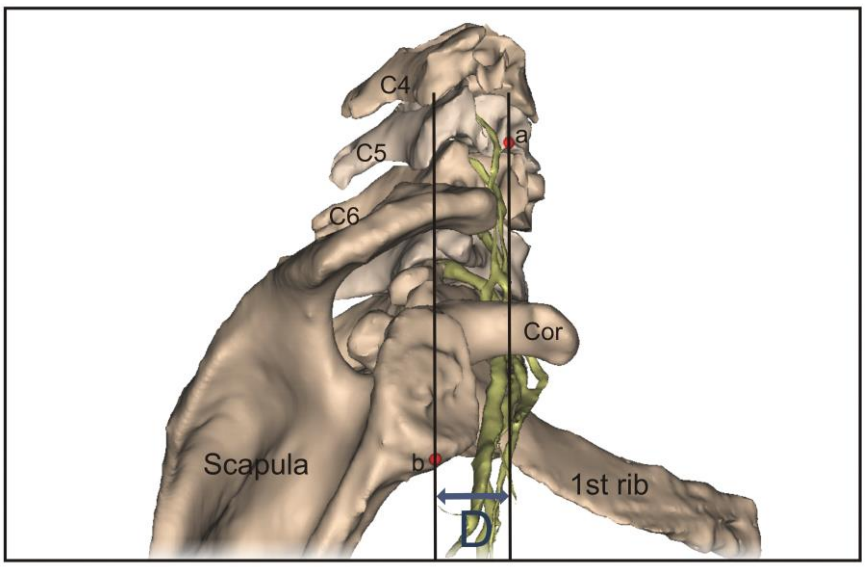


Fig. 1. Measurement of the protraction distance (D). a, anterior tubercle of C5; b, infraglenoid tubercle; Cor, coracoid process; C4, 4th cervical vertebra; C5, 5th cervical vertebra; C6, 6th cervical vertebra.

To determine the optimal number of preselected atlases, a leave-one-out strategy was followed: One of the 12 available cadaver CT-datasets was selected as a patient and the remaining CT-datasets, which contained the anatomically validated BP segmentation, served as atlases. For multi-atlas-based BP autosegmentation, the atlases in which the protraction distances were closest to the protraction distance of the patient were selected. The number of atlases selected started at 2 and gradually increased up to 8. This procedure was repeated using each atlas as a patient (Figure 2). The combinations with 9, 10 and 11 atlases were not calculated because from 6 atlases, a deceasing trend was observed and would only have continued, given that only 12 atlases were included in this study. This strategy resulted in a total of 84 (7 x 12) combinations over the different number of

atlases. A Power analysis was conducted (power π= 80) to calculate the minimum sample size needed for a 90% confidence interval.

In the second step, similarity indices were calculated measuring the similarity between each generated multi-atlas-based autosegmentation and the gold standard BP segmentation.

First, the Dice similarity coefficient (DSC) was calculated between the automatic segmentation and the gold standard. The DSC measures the spatial overlap between the gold standard A and the autosegmentation B, and it is defined as DSC(A,B) = 2(A∩B)/(A+B) where ∩ is the intersection volume. The DSC is situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement. This coefficient linearly increases with the increment of overlap between the 2 segmentations and gives a penalty for a false positive delineation area.

We also calculated the Jaccard index (JI) as the ratio of the intersection volume and the entire union volume of the delineations: JI(A,B) = (A∩B)/(AUB). The JI is also situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement. This coefficient has a non-linear increment. The penalty for a false positive delineation area increases faster compared to the DSC.

Finally, the inclusion index (INI) was measured between the gold standard BP (A) and the registered BP (B). INI is the intersection volume of both, divided by the gold standard BP: INI = (A∩B/A) [7]. INI is situated between 0 and 1, with 0 indicating no inclusion and 1 indicating total inclusion of A by B. This coefficient linearly increases with the increment of overlap between the 2 segmentations and gives no penalty for a false positive delineation area.

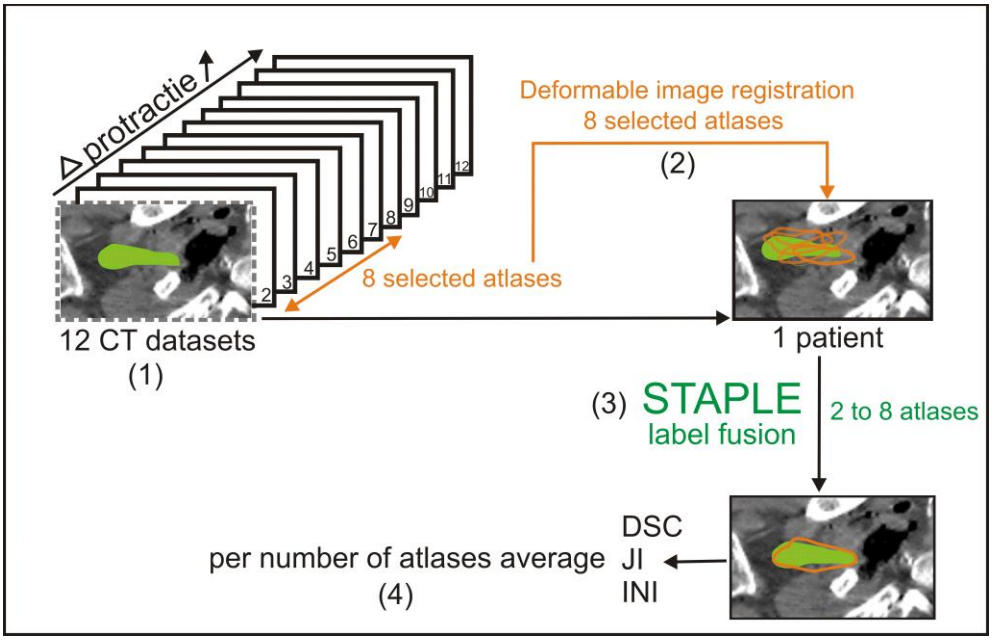


Fig. 2. Schematic illustration of the procedure for one patient. (1) 12 cadaver CT datasets were included, one atlas was selected as a patient and the 11 other atlases were morphometrically ranked relative to this patient. (2) The first 8 morphometrically ranked atlases were taken for deformable image registration on the patient. (3) Label fusion was performed with 2 up to 8 atlases. (4) Per the number of atlases, the average Dice similarity coefficient (DSC), Jaccard index (JI) and Inclusion index

(INI) were calculated for the generated contour with the gold standard contour. This procedure was repeated using every atlas as a patient.

The difference in similarity index values between the different number of atlases was assessed using an equivalence trial [13, 14], which is used to demonstrate similarity between compared groups. It uses a confidence interval where equivalence is claimed when the confidence interval of the difference in outcome between compared groups is within a predetermined equivalence margin. This equivalence margin represents a clinically acceptable range of differences. For this study, an equivalence margin of 10 percent was premised.

The number of atlases with the highest average DSC was chosen as a reference group for the equivalence trial. DSC was chosen above JI and INI because the DSC has a linear course with increasing correctly delineated volume. This means that a 10 percent (= equivalence margin) increment or decrement of this index always correlates with the same amount of increment or decrement of correctly delineated volume. With JI, on the other hand, the amount of correctly delineated volume associated with an increment or decrement of 10 percent of the JI value will vary depending on the starting value of the JI, because this index has a non-linear course. The inclusion index was not adequate for the equivalence trial because the highest INI value does not necessary imply the most accurate segmentation [3].

Starting from this reference group, the number of atlases was first gradually increased by one. If, by increasing the number of atlases each time starting from the reference group, the decrement of DSC (90% confidence interval (CI)) fell within the equivalence margin of 10 percent, then groups were considered as equivalent. Next, the number of atlases was gradually decreased by one, starting from the reference group. If, by decreasing the number of atlases each time starting from the reference group, the decrement of similarity index values fell within the equivalence margin, then the calculation time could be reduced by using a lower number of atlases without a clinically relevant loss in accuracy.

Subsequently, the similarity index values of the ‘selected’ group were compared to the similarity index values of the ‘random’ group determined by Van de velde et al. in 2015 [3]. This was performed using an independent sample t-test.

Results

The purpose of this study was to measure the effect of a morphometric atlas selection strategy on multi-atlas-based BP autosegmentation in the commercially available software package ADMIRE® and to determine the optimal number of selected atlases to use.

To measure the difference between BP autosegmentation with morphometrically selected atlases and randomly chosen atlases in an independent sample-t test, the power analysis (π= 80) resulted in a minimal sample size of 75 combinations per number of atlases needed for a 90% confidence interval.

Per number of atlases, average DSC, JI, and INI for the selected and the random group are shown in table 1 and figures 3 and 4.

Table 1. Mean Dice similarity coefficient, Jaccard index, and Inclusion index per number of atlases for the selected group and for the random group.

	Selected			Random		
Number of atlases	**DSC (SD)**	**JI (SD)**	**INI (SD)**	**DSC (SD)**	**JI (SD)**	**INI (SD)**
2	0,44 (0,159)	0,296 (0,133)	0,366 (0,161)	0,247 (0,179)	0,154 (0,131)	0,188 (0,158)
3	0,535 (0,142)	0,378 (0,134)	0,521 (0,154)	0,397 (0,184)	0,265 (0,151)	0,373 (0,187)
4	0,58 (0,120)	0,417 (0,110)	0,604 (0,127)	0,472 (0,171)	0,325 (0,147)	0,473 (0,184)
5	0,589 (0,101)	0,425 (0,094)	0,677 (0,120)	0,482 (0,153)	0,331 (0,132)	0,534 (0,166)
6	**0,598 (0,107)**	**0,434 (0,101)**	0,733 (0,120)	0,519 (0,138)	0,362 (0,128)	0,616 (0,155)
7	0,593 (0,103)	0,428 (0,099)	0,75 (0,114)	0,514 (0,129)	0,356 (0,117)	0,658 (0,147)
8	0,581 (0,095)	0,416 (0,091)	**0,767 (0,113)**	0,501 (0,120)	0,343 (0,106)	0,686 (0,143)

Abbreviations: DSC, Dice similarity coefficient; JI, Jaccard index; INI, Inclusion index; SD, standard deviation. In bold: highest similarity index values.

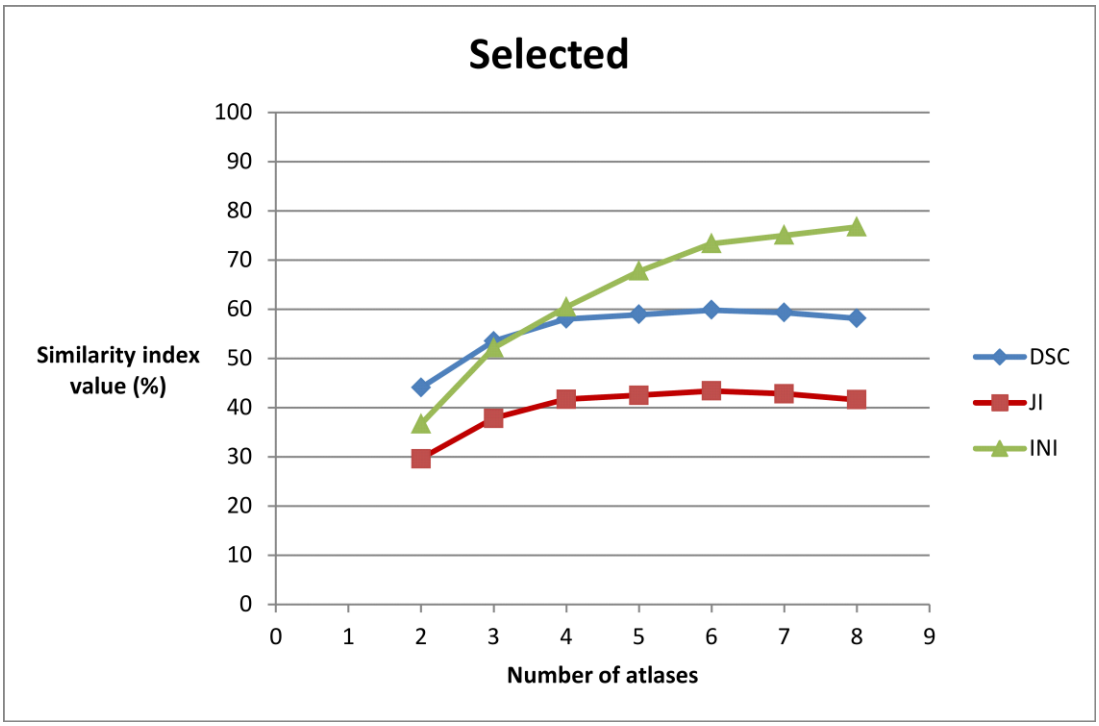


Fig. 3. The behaviour of the similarity indices with an increasing number of selected atlases. DSC, Dice similarity coefficient; JI, Jaccard index; INI, Inclusion index

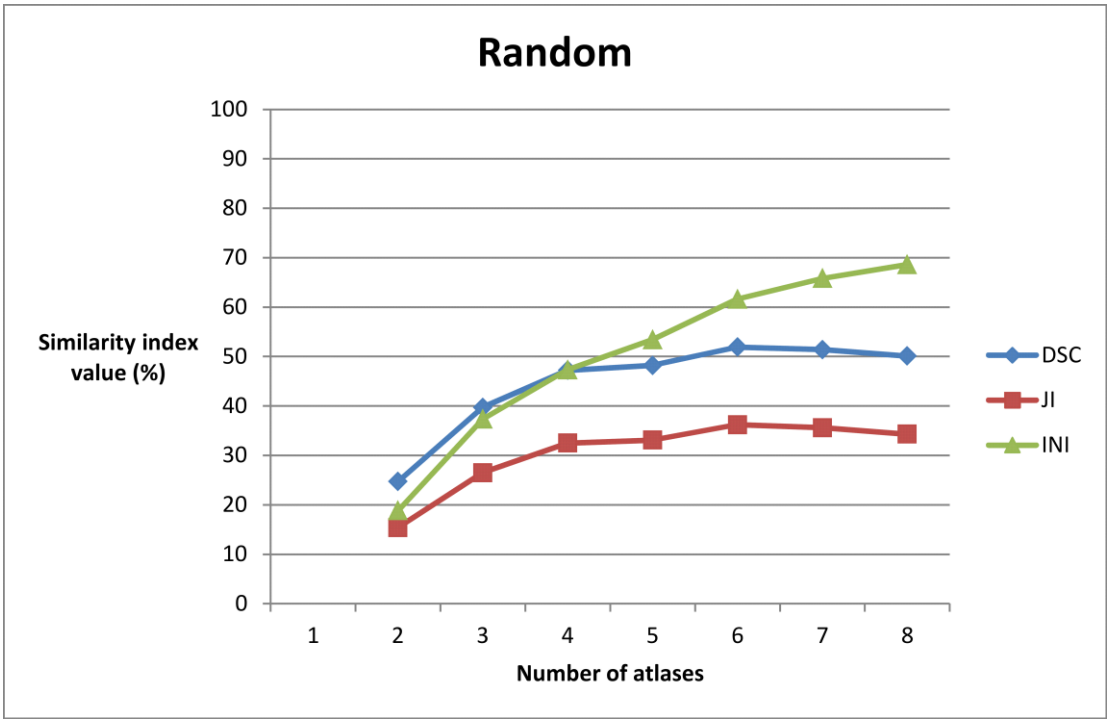


Fig. 4. The behaviour of the similarity indices with an increasing number of random atlases. DSC, Dice similarity coefficient; JI, Jaccard index; INI, Inclusion index

For each number of atlases, the difference in all average similarity index values (DSC, JI, INI) between the selected and random groups was significant ($p < 0.05$) (Table 2, Figure 5). The fewer atlases used, the bigger the difference between the random and the selected group.

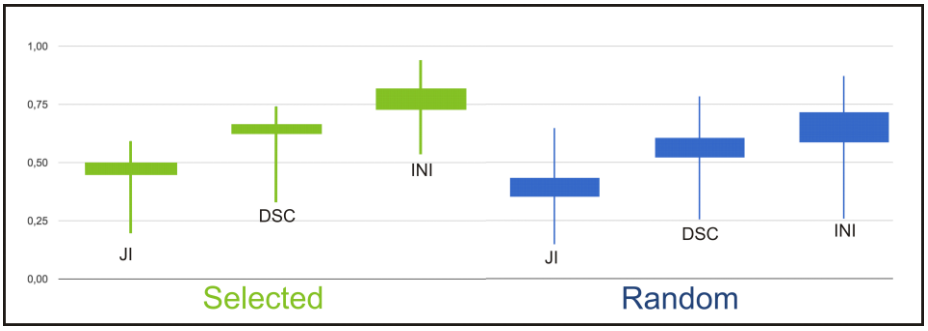


Fig. 5. Box plot showing the similarity index values in the selected group and the random atlas group for 6 atlases.

Table 2. Statistics of the differences in the mean DSC, JI, and INI values between the random and the selected groups

Number of atlases	Mean DSC random	Mean DSC selected	Mean difference	p-value
8	0,501	0,581	0,081	0,019
7	0,514	0,593	0,082	0,026
6	0,519	0,598	0,079	0,035
5	0,482	0,589	0,109	0,005
4	0,472	0,58	0,111	0,012
3	0,397	0,535	0,137	0,009
2	0,247	0,44	0,187	0,003
Number of atlases	**Mean JI random**	**Mean JI selected**	**Mean difference**	**p-value**
8	0,343	0,416	0,074	0,024
7	0,356	0,428	0,075	0,031
6	0,362	0,434	0,072	0,042
5	0,331	0,425	0,095	0,007
4	0,325	0,417	0,095	0,017
3	0,265	0,378	0,112	0,02
2	0,154	0,296	0,137	0,006

Number of atlases	Mean INI random	Mean INI selected	Mean difference	p-value
8	0,686	0,767	0,083	0,037
7	0,658	0,75	0,095	0,021
6	0,616	0,733	0,118	0,008
5	0,534	0,677	0,145	0,002
4	0,473	0,604	0,134	0,005
3	0,373	0,521	0,147	0,01
2	0,188	0,366	0,173	0,005

Abbreviations: DSC, Dice similarity coefficient; JI, Jaccard index; INI, Inclusion index.

The highest DSCs were found using 6 selected atlases (Table 1, Figure 3). By increasing the number of selected atlases from 6 to 7 atlases, the decrement of DSC values did not fall within the predisposed equivalence margin of 10 percent (p> 0.05) (Figure 6). By decreasing the number of atlases from 6 to 5 atlases, the decrease in DSC values also did not fall within the predisposed equivalence margin (p> 0.05) (Figure 6). This means that equivalence cannot be proven when using 7 or 5, instead of 6, atlases. Here, the possible number of combinations (12 combinations per number of atlases) was not large enough for a 90 percent confidence interval, a power of 80, and an equivalence margin of 10 percent.

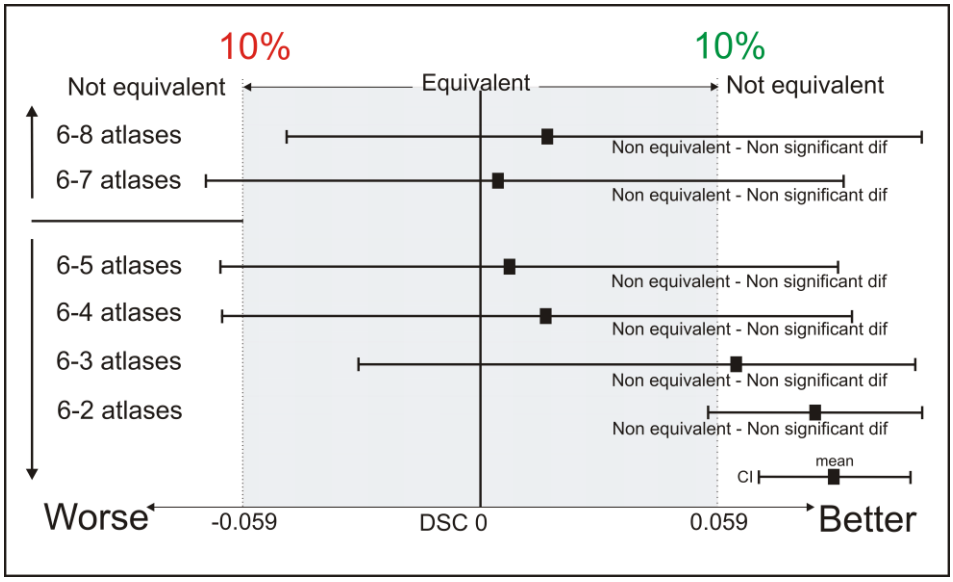


Fig. 6. Interpretation of the equivalence using 6 selected atlases in multi-atlas brachial plexus autosegmentation compared to using more (7-8) and less (5-4-3-2-1) selected atlases. The shaded area covers the equivalence range of 10%. □ = observed point estimate of outcome difference in each number of atlases, corresponding error bar= two-sided 90% confidence interval (caps at each end = lower and upper bar bounds of confidence interval). DSC, Dice similarity coefficient; CI, confidence interval.

Discussion

The current study aimed to measure the effect of morphometric atlas selection on multi-atlas-based BP segmentation and to determine the optimal number of selected atlases to use.

For each number of atlases, the difference in all average similarity index values (DSC, JI, INI) between the 'selected' and 'random' groups was significant ($p < 0.05$) (Table 2, Figure 5). The highest DSCs were found using 6 selected atlases (Table 1, Figure 3).

The results indicate that multi-atlas BP autosegmentation accuracy significantly increases when atlases were selected on the basis of the difference in protraction distance between the atlas and the patient compared to autosegmentation with random atlases. The more atlases used, the smaller the difference becomes between the random and the selected group. This can be explained by the size of the atlas database. A relatively small database means that when the number of atlases used for autosegmentation increases, fewer combinations of random atlases can be constituted, and relatively more of those combinations become similar to the combinations of the selected atlases.
To test whether the accuracy could be improved by using more than 6 atlases, similarity indices were calculated using 7 and 8 selected atlases. If, in these cases, the decrease of the DSC values compared to 6 atlases is not clinically relevant and significantly higher INI values are obtained, then this will suggest that higher accuracy is reached with 7 or 8 atlases. However, when an equivalence trial was performed to compare using 6 atlases to using 7 or 8 atlases, the decrement of the DSC's did not fall inside the equivalence margin of 10 percent (figure 6), indicating non-equivalent autosegmentation results (Figure 6). However, there was not enough statistical power to prove statistical equivalence. Therefore, more atlases need to be included to reach the sample size needed for a power of 80 and an equivalence margin of 10 percent. However, based on figures 3 and 4, it could be suggested that using 7 or 8 atlases is equivalent to using 6 atlases and that using 4 or 5 atlases is also equivalent to using 6 atlases. A reason for decreasing the number of atlas could be a reduction of the calculation time or calculation power.

The current study is the first to investigate the effect of an organ-specific atlas selection strategy. Other studies make use of organ-aspecific atlas selection strategies, such as image similarity or atlas selection, on the basis of meta-information related to the patient (like age or BMI) [4- 6]. However, these atlas selection strategies have not been applied to BP autosegmentation.

In one single publication, BP autosegmentation, specifically, was investigated [15]. In this study, the authors conclude that multi-atlas autosegmentation with 9 atlases can be effectively used to delineate the BP on CT. However, these conclusions may be unsafe because the autosegmentation itself and the validation procedure afterwards were completely based on gold standards of BP contours generated by delineators using BP contouring guidelines, which were proven to be unreliable [2]. No atlas selection strategy was applied in this study, and the authors state in their discussion that the benefit of atlas selection needs to be investigated in future studies.

In the clinic, when shoulder protraction is measured on the planning CT of a patient and the most patient-similar atlases regarding shoulder protraction position are selected out of an atlas database, autosegmentation accuracy will significantly improve compared to autosegmentation with the same amount of random atlases. This atlas selection procedure can be implemented as fully automatic. Moreover, when the protraction position of the patient can be standardized during planning-CT, protraction differences of the planning-CT and CT's in the atlas-database could be kept within bounds, which will contribute to further improvement of the autosegmentation accuracy.

A major limitation of this study is that an insufficient number of atlases was included to reach enough statistical power for comparison of the different numbers of atlases. That is why only suggestions concerning the optimal number of atlases could be made. For definitive conclusions, more atlases need to be included to cover a wider range of protraction positions and to increase statistical power. When the number of atlases is increased, the difference between multi-atlas-based autosegmentation with random atlases and selected atlases, which is clearly demonstrated in this study, will only become more distinct, and the accuracy of the autosegmentation results will also further increase due to an increasing probability of selecting atlases that are more similar to the patient's morphotype.

In the future, the dosimetric implications of morphometric atlas selection in radiotherapy planning need to be investigated. Therefore, the accuracy of the morphometric atlas selection strategy needs to be first tested on CT-datasets of head-and-neck, breast, or lung cancer patients with high-quality gold standard BP delineations included.

Conclusion

Morphometric atlas selection on the basis of the protraction position of the patient significantly improves multi-atlas-based BP autosegmentation accuracy for each number of atlases investigated in this study.

The optimal number of selected atlases to use for BP autosegmentation is 6 in this study, but for definitive conclusions about the optimal number of atlases, more than 12 atlases need to be included in the atlas database to increase statistical power.

List of abbreviations

CT, computed tomography; DSC, Dice similarity coefficient; JI, Jaccard index; INI, Inclusion index; OAR, organs at risk; BP, brachial plexus; MRI, magnetic resonance imaging; DVF, deformation vector field; STAPLE, Simultaneous Truth and Performance Level Estimation; CI, confidence interval; SD, standard deviation

Competing interests

The authors declare that they have no competing interests.

Authors contribution

JVDV and JW contributed to the conception and design, the acquisition and analysis of data, the interpretation of data, and the drafting of the manuscript. TV was involved in the acquisition of the data and the drafting of the manuscript. WDS was involved in the acquisition of the data and the drafting of the manuscript. EA contributed to the conception and design of the study, as well as the drafting of the manuscript. WDN contributed to the conception and design of the study and the drafting of the manuscript. TVH coordinated the investigation, was involved in the acquisition and analysis of data, the interpretation of data, and the drafting of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors thank Elekta AB, Stockholm, Sweden.

This research was supported by the Foundation against Cancer, grant 2012-200.

References

1. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. Med Phys. 2014;41.
2. Van de Velde J, Vercauteren T, De Gersem W, Wouters J, Vandecasteele K, Vuye P, et al. Reliability and accuracy assessment of radiation therapy oncology group-endorsed guidelines for brachial plexus contouring. Strahlenther Onkol. 2014;190:628-635.
3. Van de Velde J, Wouters J, Vercauteren T, De Gersem W, Duprez F, De Neve W, et al. Morphometric Atlas Selection for Automatic Brachial Plexus Segmentation. Int J Radiat Oncol Biol Phys. 2015;92(3):691-8.
4. Teguh DN, Levendag PC, Voet PWJ, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys. 2011;81:950-957.
5. Van Rikxoort EM, Isgum I, Arzhaeva Y, Staring M, Klein S, Viergever MA, et al. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. Medical Image Analysis. 2010;14:39-49.
6. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage. 2009;46:726-738.
7. Van de Velde J, Audenaert E, Speleers B, Vercauteren T, Mulliez T, Vandemaele P, et al. An Anatomically Validated Brachial Plexus Contouring Method for Intensity Modulated Radiation Therapy Planning. Int J Radiat Oncol Biol Phys. 2013;87(4):802-8.
8. Thiel W. Die Konservierung ganzer leichen in natürlichen farben. Ann Anat. 1992;174:185-195.
9. De Crop A, Bacher K, Van Hoof T, Smeets PV, Smet BS, Vergauwen M, et al. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. Radiology. 2012;262:298-304.
10. Warfield SK, Zou KH, Wells WM. Validation of image segmentation by estimating rater bias and variance. Philos Trans A Math Phys Eng Sci. 2008;366:2361-2375.
11. Coupe P, Manjon JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage. 2011;54:940-954.
12. Han X, Hibbard S, O'Connell N, et al. Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion. proc. MICCAI Workshop Head & Neck Autosegmentation Challenge. 2010;297–304.

13. Veronesi U, Orecchia R, Maisonneuve P, Viale G, Rotmensz N, Sangalli C, et al. Intraoperative radiotherapy versus external radiotherapy for early breast cancer (ELIOT): a randomised controlled equivalence trial. Lancet Oncology. 2013;14:1269-1277.
14. Ahn S, Park SH, Lee KH. How to Demonstrate Similarity by Using Noninferiority and Equivalence Statistical Testing in Radiology Research. Radiology. 2013;267:328-338.
15. Yang J, Amini A, Williamson R, et al. Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy. PRO. 2013;3:139-147.

4. Discussion

4. Discussion

4.1 Have the objectives of this thesis been achieved?

This doctoral thesis aimed to develop an accurate, precise and time-efficient BP segmentation method on CT for IMRT treatment planning.

1. *Development of gold standards for BP delineation.*

In a first stage, anatomically validated BP segmentations were developed as gold standards for BP segmentation on CT. The introduced method is a significant improvement compared to former BP gold standard development methods which were based on expert opinion, consensus reading or the calculation of a mathematical average contour (7, 9, 10). The latter methods are unreliable for poorly visible structures like the BP, since the absolute accuracy of these contours cannot be proven (46). The accuracy of the introduced gold standards, on the other hand, is guaranteed by the introduced process of anatomical validation, in which a large structural conformity was found between the 3D reconstructed BP segmentations and the dissections of the corresponding cadaver specimen. Even anatomical variations were observed to be equally present in the 3D reconstructions and in the dissections. Moreover, in the introduced gold standards, significantly more anatomical detail is obtained compared to former gold standards: all roots, trunks, divisions, cords and terminal branches are depicted with even visualization of most important anatomical variations. In the existing gold standards however, the BP was only depicted as one massive block without separation of the composing nerve bundles (7, 9).

There were also some differences between the gold standard segmentations and the dissected BPs that clearly related to the difference between the surface anatomy of the nerve and the internal nerve structure. MRI, being volume (voxel) based, does not well distinguish thin layers, such as the nerve sheet. The MRI revealed that BP branches may consist for some length of multiple more or less distinct bundles, which may or may not remerge into more integrated single bundles. In the 3D reconstruction these would be segmented as a nerve branching earlier or splitting for some length and possibly remerging thereafter (Figure 15). In dissection, such a nerve would still present as a single structure, as the internal nerve structure remained covered by the nerve sheath. Clearly, the thin nerve sheath lacked the cross-sectional area to be well visible in the MRI scans for reconstruction. Such nerve branch splitting artefacts in the 3D reconstructions were not rare (3/12 cadavers).

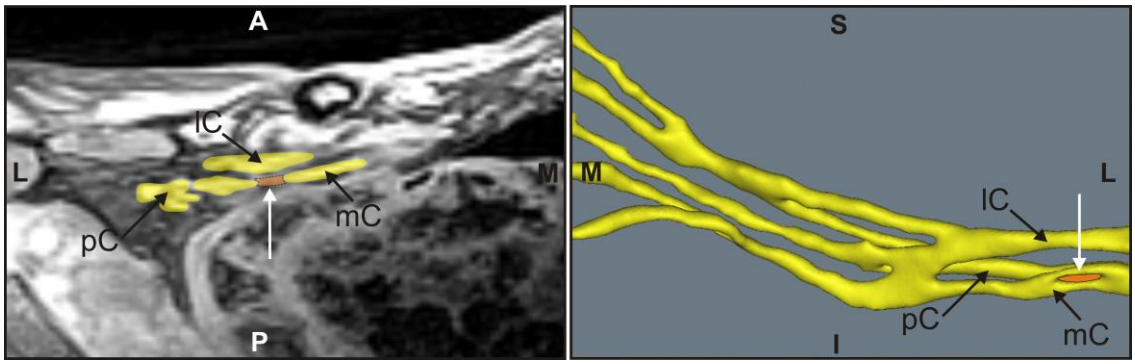


Fig. 15. Illustration of a nerve splitting artefact (white arrow). In the 3D reconstruction (posterior view) a splitting of the medial cord is visible which was not visible in the dissected BP.

We also observed that differences in the location of the nerve bifurcation can occur. It is possible that in the 3D reconstruction, the bifurcation of a nerve branch can be depicted sooner than in the dissected cadaver (mean difference of 4,4 mm). This may be due to the ability to visualize the intra-neural branching pattern of nerves into different fascicular bundles on MRI (58).

These results indicated that in MRI based reconstructions the internal nerve structure dominates and may introduce nerve branches artefact that would not be found in their macroscopic appearance during surgery.
To our knowledge, these artefacts are never described and illustrated in literature.

2. *Validation of the existing delineation guidelines for manual BP contouring.*

With the newly developed anatomical gold standards, the existing BP delineations guidelines (6) were validated in a first stage. The accuracy of the delineations following the RTOG-endorsed guidelines of Hall et al. (7) was generally found to be poor and worsened from the medial to lateral BP regions. The mean inclusion of the real BP volume was only 38 percent. Also intra- and inter-reliability were poor (mean interobserver JI: 0.038, range: 0.0-0.124; mean intraobserver JI: 0.253, range: 0.004–0.636).

Our findings are in contradiction with the reports of Hall et al. (7, 8). Both studies state that the BP contouring atlas, developed by their own group, provides a reliable set of guidelines for consistent BP contouring and a robust tool for accurate BP delineation. However, their conclusions are not substantiated by any statistical analysis. Neither accuracy nor inter- and intra-observer reliability were studied in these reports. Only visual assessment of the contours was performed without any further processing.

A second validation study for the RTOG-endorsed guidelines was published in 2014 by Min et al. (59). In this study, only the inter-observer reliability was assessed, without testing the accuracy. The authors found significant inter-observer variability related to difficulty in identifying anatomical landmarks on CT, unclear instructions in parts of the protocol and conceptual weaknesses in the guidelines published by Hall et al. A major limitation of the study of Min et al. however is that the inter-observer reliability of the guidelines of Hall et al. was also only visually assessed, together with comparison of the different BP delineation volumes but without calculation of overlap indices like JI or Apparent Volume Overlap. This means that no absolute quantification of the inter-reliability was provided.

Our study was the first accuracy assessment and the first objective quantification of the reliability of the guidelines.

3. *Development of new BP delineation guidelines for manual contouring*

New BP delineation guidelines were developed based on 8 cadaver CT-datasets with the gold standard BP delineation included. On many levels, the new guidelines are an improvement compared to the existing guidelines. First, the guidelines use unambiguous anatomic anchor points. Also, they provide more anatomic information: for each slice in the clinical dataset, a corresponding guideline exists. The introduced method of dividing the BP in 4 separate regions leads to accurate positioning of the BP in relation to appropriate reference structures, independently from the vertebral column. This anticipates for a substantial amount of errors due to anatomical variance. The combination of all these factors led to an accurate delineation protocol. The accuracy (average INI=1) improved significantly in comparison with the former guidelines (average INI= 0.38).

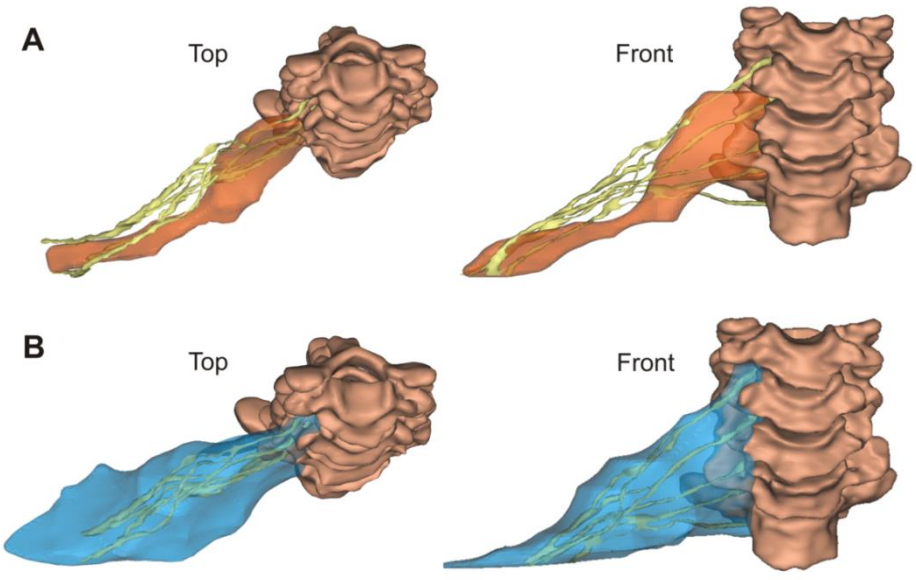


Fig. 16. Results of brachial plexus contouring in 3D following A: in orange the conventional guidelines (mean 37.75% plexus coverage) and B: in blue the new guidelines (mean 100 % plexus coverage). Left: top view, right: front view.

Hall et al. (6) were the first to introduce guidelines for manual BP contouring. These guidelines were modified by Yi et al. (7). Deficiencies of Hall's guidelines are described in the introduction (p 23). The only remaining publication concerning BP delineation guidelines specifically, was from Truong et al. (8). Here, the authors performed indirect comparison of CT and MRI to introduce 5 written steps for BP contouring. These 5 steps provide little or no relevant information, since in their guidelines the authors refer to the BP itself for BP contouring. This is a form of circular reasoning in which the initial issue is re-introduced in the solution, as the BP is invisible on CT. The only added value of this study compared to the study of Hall et al. is that slightly more visual information is provided in the form of delineated axial CT and MRI slices. In two other studies concerning BP delineation (10, 60), boundaries for BP contouring are roughly given along with delineation boundaries of several other organs. These studies were based on only 1 generated contouring atlas and existing anatomical textbook information. Herein, no substantial information for BP contouring is provided.

4. Development of a multi-atlas-based automatic BP contouring method

In an attempt to further increase the reliability and the accuracy, and to reduce the duration of the segmentation process, an automatic BP contouring method was developed. Therefore, the optimal label fusion algorithm and number of atlases had to be determined and an effective atlas selection procedure had to be developed. These topics have never been investigated for BP autosegmentation.

We only found one single study in which BP autosegmentation specifically was investigated (61). In this study, the authors concluded that multi-atlas autosegmentation with 9 atlases can be effectively used to delineate the BP on CT. However, these conclusions may be unsafe because the autosegmentation process itself and also the validation procedure afterwards were completely based on gold standards of BP contours generated by delineators using BP contouring guidelines which were proven to be unreliable (46). No gold standards based on anatomical validation were integrated, no optimal number of atlases or label fusion algorithm was investigated and no atlas selection strategy was applied in this study.

4.1 Determination of optimal label fusion algorithm and optimal number of atlases

STAPLE and Patch label fusion were compared in ADMIRE® software and the optimal number of atlases was determined using 12 cadaver CT-datasets with the anatomically validated BP delineation as a gold standard.
The results show that using 9 atlases with STAPLE label fusion results in the most accurate BP autosegmentations (average DSC= 0,532; average JI= 0,369 and average INI= 0,760). Only when using fewer than 4 atlases, the Patch label fusion results in a significantly more accurate autosegmentation than STAPLE.

The comparison of STAPLE and Patch fusion was not investigated previously for BP autosegmentation. Both label fusion algorithms were only compared for the brain stem, spinal cord, mandibular, parotis and submandibular glands in the white paper of ADMIRE® (53). The authors of the withe paper conclude that Patch fusion clearly outperforms STAPLE for all structures considered. For the BP we came largely to the opposite conclusion, which is that STAPLE mostly outperforms Patch. Still, the difference between both was very small and only significant for INI values (Manuscript 5). A reason for this small difference could be the very low BP contrast on CT, as Patch uses the CT-image intensity information for the intensity weighting. Moreover, for the entire head-and-neck, CT-image intensity information is often ambiguous and sensitive to common CT-artifacts (53). This probably makes the image information needed for assessment of the atlas registration accuracy, as performed in Patch label fusion, not sufficiently discriminative.
The current study is the first to investigate the optimal number of atlases for BP autosegmentation. The optimal number of atlases for some other organs was already studied: for the nucleus caudatus Aljabar et al. (2009) (63) concluded that using 8 atlases is optimal; for the hippocampus, the highest accuracy is reached with a selection of 25 atlases. Pirozzi et al. (2012) concluded that for the bladder and the femur, the optimal number of atlases was 5, and that the optimal number for the prostate and rectum was 4 (63). Remarkable is that in the first study (62), the number of atlases for autosegmentation of anatomically stable brain structures is higher than in the second study (63), in which anatomically variable organs were autosegmented. The opposite could be expected. The varying results of these studies only show that the optimal number of atlases is very organ-dependent and especially algorithm-dependent. So, for more general conclusions concerning the optimal number of atlases for BP autosegmentation, this has to be investigated also using other algorithms.

Few studies were found that compared ADMIRE® software to other autosegmentation software. Simmat et al. (74) found higher flexibility and robustness in the algorithm used in the ADMIRE® software compared with the algorithms in Iplan® for the bladder, prostate and rectum. La Macchia et al. (75) found the best label fusion results using STAPLE in ADMIRE® compared with the algorithms in

VelocityA® and MIM 5®, for the head-and-neck region. BP autosegmentation was not included in both studies. For general conclusions concerning the best autosegmentation software for BP autosegmentation, different autosegmentation software need to be compared in further studies.

4.2 Development of an atlas selection procedure for multi-atlas-based BP contouring

Multi-atlas-based (MAB) autosegmentation makes use of a database of multiple presegmented image datasets (called 'atlases') in order to achieve a new de novo BP delineation on a patient. To select the most patient-similar atlases out of an atlas database , a morphometric atlas selection strategy for MAB BP segmentation was introduced.

The influence of several CT-measurable morphometric parameters on BP autosegmentation accuracy was tested in order to define the best morphometric parameters for atlas selection.

Based on all of the investigated parameters, shoulder protraction distance D (Figure 13) was the most relevant parameter influencing the autosegmentation accuracy and thus preferable to select atlases for MAB BP autosegmentation. The data of this study stress the importance of a stable scapular protraction positioning protocol to minimize the positional differences between atlas and patient.

Several methods for selecting the most patient-similar atlases have been published (62, 64, 65). Most of these methods are based on the similarity between the atlas and the target image. Atlas selection strategies based on image similarities has a number of drawbacks. When an insufficient similarity exists between the atlas and the target image, local mismatches often occur, which in turn leads to segmentation errors (66-68).

Within similarity-based atlas selection strategies, a further distinction can be made between methods that use registration of the whole image (64) and those that use only part of the image (55, 69). When only a part of the image is used, it remains questionable which part should be selected. Theoretically, the best option is to use only image regions that are topographically stable in relation to the structure that has to be segmented (62).

An alternative technique of atlas selection is through the use of meta-information related to the subject. The potential selection criteria are the body mass index, age, pathology, clinical history, gender and handedness (62). For example, age-based selection has been shown to be as effective as selection based on image similarity after affine transformation (67). One major disadvantage of using meta-information is that it is not suitable when dealing with the anatomical variability that occurs independently of the simple meta-information (62).

The weakness of image similarity selection methods that utilize largely irrelevant areas of anatomical information and also the weakness of meta-information selection methods that do not address the anatomical variability, was both countered through the use of only target-specific stable anatomical information in the morphometric atlas selection process. A target-specific atlas selection method for automatic BP segmentation has to our knowledge never been published.

4.3 Testing the effect of morphometric atlas selection on multi-atlas based autosegmentation

In a final experiment, MAB autosegmentation with morphometrically selected atlases was compared to MAB autosegmentation with random atlases and also the optimal number of selected atlases was determined. For every number of atlases, average similarity indices of the morphometrically selected atlas group turned out to be significantly higher than the random group ($p<0.05$). In this study, the highest similarity indices were achieved using MAB autosegmentation with 6 selected atlases (average DSC= 0,598; average JI= 0,434; average INI= 0,733; manuscript 6). But, concerning the optimal number of atlases, no definitive conclusions can be drawn due to a lack of statistical power. To increase the statistical power and to draw definitive conclusions, more atlases need to be included in the study. The more atlases included in the study, the more accurate the autosegmentation results will be as well, because the probability of selecting atlases that are more similar to the patient's morphotype will increase.

4.2 Future directions

4.2.1 Accuracy testing of BP autosegmentations on patients

In this thesis, anatomically validated gold standard BP segmentations on CT were developed using cadaver imaging datasets. By implementing these CT's and gold standards as atlases for MAB autosegmentation on patients, we already succeeded to generate plausible de novo BP segmentations on patients too. However, it was not yet possible to measure the absolute accuracy of these segmentations due to the absence of a qualitative gold standard on patients. Therefore, in a next research step, gold standards need to be developed on patients to adequately measure the accuracy of the autosegmentation method. To this end, we already managed to produce accurate BP segmentations on MRI's of head-and-neck cancer patients. However, to this point, direct MRI-CT fusions are not yet optimized in vivo. The problem here is that the patient positioning between CT and MRI imaging is hard to standardize. Until now, we are optimizing the procedure to obtain MRI and CT of the same patient in a perfectly standardized position and an identical moment in treatment. Imaging at an identical moment in treatment is necessary because otherwise, alterations

in tumor and soft tissue volumes occur. Due to this positioning difference and the changed tumor and soft tissue volumes, rigid image fusion algorithms do not suffice to fuse these images. Deformable image fusion algorithms are needed for these purposes. However, the testing of different deformable CT-MRI fusion methods did not gave satisfactory results until now. As long as no optimal deformable fusion method is found, an adequate assessment of the accuracy of our autosegmentation method on patients stays difficult.

4.2.2 BP autosegmentation in other treatment positions

The BP delineation guidelines and the automatic BP contouring method proposed in this thesis are only investigated for patients positioned in supine with both arms alongside the body in neutral position. This is the standard treatment position for patients with head-and-neck tumors. However, for patients with lung top tumors, the arms are elevated above the head, and for breast tumor patients, there is a trend toward prone positioning. The guidelines and the BP autosegmentation method is not investigated for those specific treatment positions. Separate atlases need to be developed to come to satisfactorily results for BP autosegmentation in other treatment positions.

4.2.3 Investigation of the dosimetric implications of BP autosegmentation

Additional studies are necessary to investigate the dosimetric impact of the application of the newly developed BP autosegmentation method on IMRT planning and to measure the potential benefit regarding tumor irradiation and sparing of OARs for patients undergoing radiation therapy treatment. Also, clinical trials have to be conducted in which the incidence of brachial plexopathy in patients having the old BP delineation guidelines, can be directly compared with the incidence in patients having our new autosegmentation method.

4.2.4 Further improvement of BP autosegmentation accuracy

To further improve the accuracy and precision of the BP autosegmentation and to increase statistical power during accuracy testing, additional atlases need to be included and the effectiveness of other atlas selection strategies can be investigated. Our future plan is to develop a more advanced atlas selection strategy by creating groups of BP atlas morphotypes making use of statistical shape analysis. For BP autosegmentation, only BP-related anatomy can be integrated in the shape model. A new patient can then be classified in one of the predefined BP morphotype atlas groups for highly detailed atlas selection.

Another plan is to replace the deformable fusion algorithm as used in Admire®, which makes use of the entire image for creating a deformation of the BP segmentation, by a more specific algorithm for

BP autosegmentation. This algorithm will be based on an deformable transformation of only BP-related stable anatomical landmark points. These BP-related anatomical anchor points could be automatically detected on CT and transformed to the patient-CT coordinate system. With the resulting deformation vector field, the BP delineation of the presegmented image data set can be deformed on the patient image data set. Since all redundant image information with a variable relationship related to the BP is omitted in this method, we expect that the results obtained with this new algorithm will be more accurate than those obtained with BP autosegmentation using convenient deformable image registration algorithms.

4.2.5 Application of the same methodology on other anatomical regions

The methodology described in this doctoral thesis can also be applied to develop autosegmentations of other anatomical regions. Axillary lymph node region delineation for example, is of extreme interest for patients with breast cancer, since malignant spread to axillary lymph nodes is one of the most important predictors of survival in breast cancer patients, which is the leading cause of cancer mortality in women worldwide (72). First attempts to develop gold standards for axillary lymph vessel delineation on CT using cadavers were already made by our research team. Retrograde lymph vessel injection with radiopaque contrast agent is the best approach to avoid that branches are missed. Lymphatic valves however, impede retrograde flow under normal conditions. Several methods of retrograde lymph vessel injections through the thoracic duct with different contrast agents were performed under various pressures in order to diffract the lymphatic valves. Until now, we only succeeded to inject the thoracic duct retrogradely. When also smaller lymph vessels can be injected and visualized on CT or MRI, gold standard delineations of axillary lymph vessels and subsequently a lymph region autosegmentation method on CT can be developed.

Autosegmentations for the lumbosacral plexus (LSP) can also be developed, using an almost identical approach as described in this thesis. Radiation-induced lumbosacral plexopathy (RILSP) has been described in the literature for patients having undergone radiotherapy for pelvic malignancies. Most cases are described in patients treated with radiotherapy for cervical cancer. However, reports of RILSP in patients treated for lower gastrointestinal malignancies and prostate cancers, have also been reported (Sun et al. 2011). Our first attempts to visualize the LSP on CT and MRI of Thiel cadaver were promising. The major branches could already be identified on MRI, but smaller details of the LSP were not yet visualized. Imaging of fresh frozen cadaver specimen could provide the solution to obtain even more contrast for LSP visualization than Thiel cadavers.

4.3 Final conclusion

In this thesis, a procedure for developing accurate gold standard BP delineations on CT was proposed. This gold standard was based on fused CT-MRI imaging of cadavers and was anatomically validated by dissection (Van de Velde 2015).

Based on these gold standards, BP delineation guidelines for manual contouring on CT were developed resulting in highly accurate BP delineations (Van de Velde 2013).

Subsequently, to reduce the time-expenditure needed for BP delineation, a multi-atlas-based automatic BP segmentation procedure was developed.

Using six morphometrically selected CT-atlases, with anatomically validated BP delineations fused by the STAPLE label fusion algorithm ultimately gave the best average BP autosegmentations result (average DSC= 0,598; average JI= 0,434; average INI= 0,733). This means that accuracy was almost doubled compared to the existing manual delineations methods (average INI= 0,38). Calculation time to accomplish these autosegmentations was 15 to 20 minutes on average (Van de Velde 2014, Van de Velde 2015).

In future, the accuracy and precision of BP autosegmentation need to be improved by using a larger atlas database, fine-tuning the atlas selection strategy and testing other registration algorithms.

5. Nederlandse samenvatting

5. Nederlandse samenvatting

Door de evolutie van conventionele radiotherapiemethoden naar de geavanceerde Intensity modulated radiotherapy (IMRT) in de recente radiotherapiekliniek is het accuraat deliniëren van tumor en risico-organen op plannings-CT van steeds toenemend belang. In die mate zelfs dat het deliniëren van anatomische structuren de zwakste schakel is geworden binnen de radiotherapie planningsprocedure. De accuraatheid van de recentste behandelingstechnieken is namelijk groter dan de accuraatheid van de anatomische delineaties waar de technieken zich op baseren. Dit geldt zeker voor organen die moeilijk of niet zichtbaar zijn op de plannings-CT, zoals de plexus brachialis (PB). Accurate PB delineaties zijn in het bijzonder moeilijk uit te voeren gezien de complexiteit van dit orgaan en zijn slechte zichtbaarheid op plannings-CT.

Een aantal factoren zorgen er bovendien voor dat de delineatie van de PB als risico-orgaan aan belang wint. Vooreerst bleken door radiotherapie geïnduceerde brachial plexopathies (RIBP) in het verleden sowieso ondergerapporteerd. Recente studies tonen aan dat zowel bij patiënten met hoofd-hals- en longtopkanker als bij patiënten met borstkanker RIBP een hogere incidentie heeft dan eerst werd aangenomen. Bovendien werd aangetoond dat door nieuwe evoluties binnen de radiotherapie, zoals IMRT en hypofractonatie, de incidentie hoger ligt dan voorheen, wanneer de PB niet correct als risico-orgaan wordt gedelinieerd. Hierbij komt nog dat kankerpatiënten steeds vaker en steeds langer overleven, waardoor het steeds belangrijker wordt om complicaties van de radiotherapiebehandeling zoveel mogelijk te vermijden en waardoor ook langetermijncomplicaties die zich vroeger niet voordeden, kans hebben zich te manifesteren.

Daarom was het doel van dit doctoraat het ontwikkelen van een accurate, precieze en tijdsefficiënte PB-segmentatiemethode op CT.

Om dit doel te bereiken was het in een eerste fase noodzakelijk om CT-datasets te ontwikkelen waar een hoogkwalitatieve gouden standaard PB-delineatie was op aangeduid. Deze gouden standaarden zijn noodzakelijk als vertrekpunt voor het creëren van algemene richtlijnen voor manuele PB-delineatie. In het bestaande PB-delineatie-onderzoek zijn deze gouden standaarden meestal gebaseerd op de mening van een groep experten die een soort consensus overeenkomen over de ligging van een orgaan. Voor de PB is dit echter een onbetrouwbare methode om gouden standaarden te ontwikkelen, gezien de complexiteit van het orgaan en zijn onzichtbaarheid op CT. Daarom wordt in deze doctoraatsthesis een nieuwe methode voorgesteld om gouden standaard segmentaties op CT te ontwikkelen. Deze gouden standaarden worden gecreëerd door middel van gefuseerde CT- en MRI-beelden van kadavers. De beelden worden vervolgens in 3D gereconstrueerd

en anatomisch gevalideerd door dissectie van het kadaver. De conformiteit van de 3D-reconstructies en de dissecties post-hoc bepaalt uiteindelijk de accuraatheid van de ontwikkelde gouden standaard. Op deze manier werden 15 CT-datasets ontwikkeld waarop een anatomisch gevalideerde gouden standaard PB-delineatie is aangeduid.

Na het ontwikkelen van de gouden standaarden werd eerst de accuraatheid en betrouwbaarheid van de bestaande PB-delineatierichtlijnen gepubliceerd door Hall et al. in 2008, onderzocht. De accuraatheid en inter- en intra betrouwbaarheid werd onderzocht door middel van delineaties van 5 verschillende delineatoren die 3 keer 3 CT-datasets inkleurden volgens de richtlijnen van Hall et al. Deze inkleuringen werden vergeleken met de ontwikkelde gouden standaarden. De accuraatheid van deze delineaties bleek zeer laag. Gemiddeld coverde slechts 38 percent van het delineatievolume de gouden standaard. Dit wil zeggen dat 62 percent van het werkelijke PB-volume ongecoverd bleef. Ook de inter- en intrabetrouwbaarheid van de delineaties was laag. Zowel de accuraatheid als de betrouwbaarheid daalden naarmate meer lateraal gelegen PB-zones werden ingekleurd.

In een tweede fase werden nieuwe richtlijnen opgesteld voor manuele PB-delineatie op CT. Zeven kadaver CT-datasets met de gouden standaard PB-delineatie dienden hiervoor als uitganspunt. De PB werd onderverdeeld in 4 onafhankelijke delineatiezones, waarbij per zone telkens verschillende CT-referentiepunten werden gebruikt voor de delineatie. De gekozen referentiepunten waren nauw verweven met de plexus en hadden een min of meer stabiele ligging t.o.v. dit orgaan. De onderverdeling van de PB in 4 onafhankelijke delineatiezones en het kiezen van zone-specifieke referentiepunten die duidelijk CT-definieerbaar zijn, zorgden ervoor dat veel accuratere inkleuringen werd bekomen.

Om de tijdsintensiviteit van de manuele segmentaties te verminderen en om te trachten de vals positieve delineatiezone te verkleinen, werd in een derde fase van het onderzoek gestart met het ontwikkelen van een automatische PB-inkleuringsmethode.

Een multi-atlas-based methode werd gehanteerd waarbij op basis van een aantal vooraf gesegmenteerde template CT-datasets (‘atlassen’) een delineatie op een nieuwe ongesegmenteerde CT-dataset (‘patiënt’) wordt berekend. Bij deze methode moet op 3 belangrijke vragen een antwoord worden gevonden om tot een optimaal resultaat te komen:

i) Welke label fusie algoritme moet worden gebruikt?, ii) Hoeveel atlassen moeten worden gebruikt? en iii) Welke atlassen moeten worden gebruikt?

Om een antwoord te vinden op vraag i) en ii) werd een leave-one-out procedure gevolgd:

Uit een database van 12 beschikbare kadaver-CT's met de gouden standaard-PB inbegrepen werd telkens 1 CT geselecteerd als 'patiënt' en de overige 11 CT's werden gebruikt als atlassen voor multi-atlas-based PB automatische segmentatie in ADMIRE® software. Het aantal gebruikte atlassen startte met 2 oplopend tot 11, met telkens alle mogelijke combinaties van atlassen. De accuraatheid van de autosegmentaties werd getest door de resultaten te vergelijken met de gouden standaard inkleuringen. Deze procedure werd 2 maal uitgevoerd: 1 maal met het STAPLE algoritme en een tweede maal met het Patch algoritme in ADMIRE®. Uit deze procedure werd berekend hoeveel het optimaal aantal atlassen is en welk label fusie algoritme het beste resultaat geeft voor PB autosegmentatie. Negen atlassen met het STAPLE resulteerde in de meest accurate autosegmentaties: gemiddeld werd 76 percent van de gouden standaard gecoverd door het delineatievolume. Tot zover betekent dit dus een verdubbeling t.o.v. de gemiddelde covering van de delineaties waarbij de richtlijnen van Hall et al. werden toegepast.

Om voor klinische toepassingen de accuraatheid van de autosegmentaties nog te verbeteren kunnen bijkomend de beste atlassen worden geselecteerd voor autosegmentatie. Dit brengt ons bij vraag iii): welke atlassen moeten worden geselecteerd voor een optimaal resultaat? Voor de selectie van de meest geschikte atlassen werd in deze doctoraatsscriptie een nieuwe methode ontwikkeld: 'Morfometrische atlas selectie'. Bij deze methode worden op elke atlas- en patiënt-CT eerst een aantal op CT zichtbare punten met een stabiele ligging t.o.v. de PB aangeduid. Vervolgens worden automatisch een aantal metingen uitgevoerd van afstanden, verhoudingen van afstanden en oppervlakten tussen deze punten. Nadien wordt gekeken welke van deze metingen een invloed heeft op de accuraatheid van de autosegmentaties. De metingen met het meeste invloed zullen worden gebruikt om atlassen te selecteren voor multi-atlas-based segmentatie. Eén meting kwam naar voren als meest beïnvloedende factor voor PB-autosegmentatie: de meting die de protractiepositie van de patiënt definieert. Deze protractie-afstand werd dan ook als criterium gebruikt om atlassen te selecteren voor PB-autosegmentatie.

In een laatste studie wordt onderzocht of het selecteren van een groep atlassen door middel van morfometrische atlasselectie een significant beter resultaat oplevert vergeleken met een willekeurig gekozen groep atlassen en of het optimaal aantal geselecteerde atlassen hetzelfde is als het optimaal aantal niet-geselecteerde atlassen. Uit deze studie blijkt dat PB multi-atlas-based autosegmentatie met morphometrische atlasselectie op basis van de protractie-positie van de patiënt, significant accuratere resultaten geeft dan multi-atlas-based autosegmentation met willekeurige atlassen en wordt de suggestie gemaakt dat met zes atlassen het optimale resultaat wordt bereikt.

6. Summary

6. Summary

Due to the recent evolution from conventional radiotherapy treatment to the advanced Intensity Modulated Radiotherapy treatment (IMRT), accurate delineation of tumor and organs at risk (OARs) has become increasingly important in radiotherapy clinic. Even to the extent that the delineation of anatomical structures is representing the weakest link in the radiotherapy planning procedure. This certainly applies to poorly visible structures on planning CT. One of these poorly visible OARs for radiotherapy planning of head-and-neck, lung and breast cancer patients is the brachial plexus (BP). Accurate BP delineations are especially difficult to achieve due to the complexity of this organ and his invisibility on planning CT.

Therefore, the goal of this doctoral thesis was to develop an accurate, precise and time-efficient BP segmentation method on CT for IMRT treatment planning.

To achieve this goal, it was necessary to develop CT-datasets with a high-quality BP gold standard delineation indicated in an initial stage. These gold standard delineations were an essential starting point for manual BP delineation guideline development. To this end, existing gold standard BP delineations in BP delineation research were based on expert opinions. However, this method is unreliable for the BP, because of the complexity of this organ and his nearly invisibility on planning CT. For this reason, a new method for the development of gold standard BP delineations on CT was introduced in this thesis. These gold standards were created by direct CT-MRI fusion of cadaver imaging datasets. Subsequently, these fused images were reconstructed in 3D and anatomically validated by dissection. Conformity of the 3D reconstructions and the corresponding dissections eventually determined the accuracy of the gold standard. As such, 15 CT-datasets were developed with the anatomically validated gold standard BP delineation included.

After the development of the gold standards, the reliability and accuracy of the existing BP delineation guidelines, published by Hall et al. in 2008, were investigated. Five different delineators delineated the BP 3 times, on 3 different cadaver CT datasets. These delineations were compared with the developed gold standard delineations. The results of this study showed a poor accuracy and reliability of the guidelines. On average, the delineations covered only 38 percent of the gold standard, and the accuracy clearly decreased from medially to laterally situated BP regions.

In a second stage, new BP delineation guidelines for manual delineation on CT were developed, making use of 7 cadaver CT datasets with the gold standard BP delineation included. Therefore, the BP was divided in 4 independent delineation regions, each with their own distinct reference points

which had a stable localization with regard to the BP. The division of the BP in 4 independent delineation regions and the allocation of region-specific reference points eventually resulted in more accurate delineation guidelines.

However, an important disadvantage of the developed delineation guidelines was their time-intensiveness. To reduce the time-intensiveness of the manual segmentations, an automatic BP segmentation method was initiated in a third stage.

A multi-atlas-based method was adopted, in which a de novo BP segmentation is calculated making use of a number of presegmented template CT datasets (called 'atlases'). To achieve optimal results, three important questions need to be solved concerning a multi-atlas-based autosegmentation procedure: i) Which label fusion algorithm has to be used?, ii) How many atlases need to be used? and iii) Which atlases have to be used?

To solve the first and the second question, a leave-one-out procedure was followed. Out of 12 atlases, one atlas was selected as a patient and the 11 remaining atlases were registered onto this patient using a deformable image registration algorithm. Next, label fusion was performed by using every possible combination of 2 to 11 atlases, using 2 different label fusion algorithms. This procedure was repeated for every atlas as a patient. It led to the determination of the optimal number of atlases and the optimal label fusion algorithm for BP autosegmentation. Nine atlases with the STAPLE algorithm resulted in the most accurate autosegmentations: an average coverage of 76 percent was achieved using this parameters.

To further increase the accuracy of the autosegmentations for clinical purposes, the best atlases needed to be selected. This led us to question iii) 'Which atlases need to be selected for an optimal result?'. For this purpose, a new atlas selection procedure was introduced in this thesis: 'morphometric atlas selection'.

This method started with the determination of several corresponding CT-visible bony points on the patient-CT and the atlases. Subsequently, a number of automatic measurements were performed based on the localization of these points. Afterwards, the influence of these measurements on BP autosegmentation was determined. One measurement appeared to be the most relevant influencing factor of BP autosegmentation: the shoulder protraction distance. As such, this shoulder protraction distance was used for atlas selection for BP autosegmentation.

In a final study, BP autosegmentation with a group of morphometrically selected atlases was compared to BP autosegmentation with a group of randomly chosen atlases. This study showed that multi-atlas-based autosegmentation with a group of morphometrically selected atlases gives significantly better results than with a group of randomly chosen atlases.

The studies conducted in this thesis leaded to BP autosegmentations with doubled accuracy compared to manual BP delineations following the existing delineation guidelines. However, for clinical application, the accuracy of the BP autosegmentation procedure introduced in this thesis needs to be further improved by expanding the atlas database and refining the atlas selection procedures and deformable fusion algorithm.

7. References

7. References

1. Halperin E, Perez C, Brady L. Preface to the first edition. Perez and Brady's principles and practice of radiation oncology, 5th ed. Philadelphia, PA: Lippincott Williams and Wilkins, 2008: xxi.
2. Galvin JM, De Neve W. Intensity modulating and other radiation therapy devices for dose painting. *J Clin Oncol* 2007;25:924-930.
3. Nelms BE, Tome WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: Test case from a patient with oropharyngeal cancer. *International Journal of Radiation Oncology Biology Physics* 2012;82:368-378.
4. Allozi R, Li XA, White J, Apte A, Tai A, Michalski JM, Bosch WR, El Naqa I. Tools for consensus analysis of experts' contours for radiotherapy structure definitions. *Radiother Oncol* 2010;97:572-578.
5. Bortfeld T, Jeraj R. The physical basis and future of radiation therapy. *British Journal of Radiology* 2011;84:485-498.
6. Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. *J Med Imag Radiat Oncol* 2010;54:401-410.
7. Hall WH, Guiou M, Lee NY, Dublin A, Narayan S, Vijayakumar S, Purdy JA, Chen AM. Development and validation of a standardized method for contouring the brachial plexus: Preliminary dosimetric analysis among patients treated with imrt for head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics* 2008;72:1362-1367.
8. Yi SK, Hall WH, Mathai M, Dublin AB, Gupta V, Purdy JA, Chen AM. Validating the rtog-endorsed brachial plexus contouring atlas: An evaluation of reproducibility among patients treated by intensity-modulated radiotherapy for head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics* 2012;82:1060-1064.
9. Truong MT, Nadgir RN, Hirsch AE, Subramaniam RM, Wang JW, Wu R, Khandekar M, Nawaz AO, Sakai O. Brachial plexus contouring with ct and mr imaging in radiation therapy planning for head and neck cancer. *Radiographics* 2010;30:1095-1103.
10. Kong FM, Ritter T, Quint DJ, Senan S, Gaspar LE, Komaki RU, Hurkmans CW, Timmerman R, Bezjak A, Bradley JD, Movsas B, Marsh L, Okunieff P, Choy H, Curran WJ. Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus. *International Journal of Radiation Oncology Biology Physics* 2011;81:1442-1457.

11. Feng GS, Lu HM, Liang Y, Chen HS, Shu LY, Lu S, Zhu JF, Gao WW. Radiation dose to the brachial plexus in nasopharyngeal carcinoma treated with intensity-modulated radiation therapy: An increased risk of an excessive dose to the brachial plexus adjacent to gross nodal disease. *Experimental and Therapeutic Medicine* 2012;4:216-220.
12. Small W, Jr., Mell LK, Anderson P, Creutzberg C, De Los Santos J, Gaffney D, Jhingran A, Portelance L, Schefter T, Iyer R, Varia M, Winter K, Mundt AJ. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy in postoperative treatment of endometrial and cervical cancer. *International Journal of Radiation Oncology Biology Physics* 2008;71:428-434.
13. Chen AM, Hall WH, Li BQ, Guiou M, Wright C, Mathai M, et al. Intensity-modulated radiotherapy increases dose to the brachial plexus compared with conventional radiotherapy for head and neck cancer. *British Journal of Radiology*. 2011;84(997):58-63.
14. Johansson S, Svensson H, Denekamp J. Dose response and latency for radiation-induced fibrosis, edema, and neuropathy in breast cancer patients. *International Journal of Radiation Oncology Biology Physics*. 2002;52(5):1207-19.
15. Delanian S, Lefaix J-L, Pradat P-F. Radiation-induced neuropathy in cancer survivors. *Radiother Oncol* 2012;105:273-282.
16. Galecki J, Hicer-Grzenkowicz J, Grudzien-Kowalska M, Michalska T, Zalucki W. Radiation-induced brachial plexopathy and hypofractionated regimens in adjuvant irradiation of patients with breast cancer - a review. *Acta Oncologica* 2006;45:280-284.
17. Amini A, Yang JZ, Williamson R, McBurney ML, Erasmus J, Allen PK, Karhade M, Komaki R, Liao ZX, Gomez D, Cox J, Dong L, Welsh J. Dose constraints to prevent radiation-induced brachial plexopathy in patients treated for lung cancer. *International Journal of Radiation Oncology Biology Physics* 2012;82:E391-E398.
18. Chen A, Wang PC, Daly ME, Cui J, Hall WH, Vijayakumar S, Phillips TL, Farwell DG, Purdy JA. Dose--volume modeling of brachial plexus-associated neuropathy after radiation therapy for head-and-neck cancer: findings from a prospective screening protocol. *Int J Radiat Oncol Biol Phy* 2014;15;88(4):771-7
19. Fathers E, Thrush D, Huson SM, Norman A. Radiation-induced brachial plexopathy in women treated for carcinoma of the breast. *Clin Rehabil* 2002;16:160-165.
20. Mutter RW, Lok BH, Dutta PR, Riaz N, Setton J, Berry SL, Goenka A, Zhang Z, Rao SS, Wolden SL, Lee NY Constraining the brachial plexus does not compromise regional control in oropharyngeal carcinoma. *Radiat Oncol*. 2013;9;8:173.

21. Chen AM, Hall WH, Li J, Beckett L, Farwell DG, Lau DH, Purdy JA. Brachial plexus-associated neuropathy after high-dose radiation therapy for head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics* 2012;84:165-169.
22. Lundstedt D, Gustafsson M, Steineck G, Alsadius D, Sundberg A, Wilderang U, Holmberg E, Johansson KA, Karlsson P. Long-term symptoms after radiotherapy of supraclavicular lymph nodes in breast cancer patients. *Radiother Oncol* 2012;103:155-160.
23. Schierle C, Winograd JM. Radiation-induced brachial plexopathy: Review. Complication without a cure. *Journal of Reconstructive Microsurgery* 2004;20:149-152.
24. Emami B, Lyman J, Brown A, Coia L, Goitein M, Munzenrider JE, Shank B, Solin LJ, Wesson M. Tolerance of normal tissue to therapeutic irradiation. *International Journal of Radiation Oncology Biology Physics* 1991;21:109-122.
25. Johansson S, Svensson H, Denekamp J. Dose response and latency for radiation-induced fibrosis, edema, and neuropathy in breast cancer patients. *International Journal of Radiation Oncology Biology Physics* 2002;52:1207-1219.
26. Chatterjee S, Lee D, Kent N, Wintle T, Mott JH, Kelly CG, et al. Managing Supraclavicular Disease from Breast Cancer with Brachial Plexus-sparing Techniques using Helical Tomotherapy. *Clin Oncol*. 2011;23(2):101-7.
27. Vargas MI, Viallon M, Nguyen D, Beaulieu JY, Delavelle J, Becker M. New approaches in imaging of the brachial plexus. *Eur J Radiol* 2010;74:403-410.
28. Viallon M, Vargas MI, Jlassi H, Lovblad KO, Delavelle J. High-resolution and functional magnetic resonance imaging of the brachial plexus using an isotropic 3d t2 stir (short term inversion recovery) space sequence and diffusion tensor imaging. *Eur Radiol* 2008;18:1018-1023.
29. Lutz AM, Gold G, Beaulieu C. Mr imaging of the brachial plexus. *Neuroimaging Clin N Am* 2014;24:91.
30. Andreou A, Sohaib A, Collins DJ, Takahara T, Kwee TC, Leach MO, et al. Diffusion-weighted MR neurography for the assessment of brachial plexopathy in oncological practice. *Cancer Imaging.* 2015;15.
31. Vargas MI, Viallon M, Nguyen D, Delavelle J, Becker M. Diffusion tensor imaging (dti) and tractography of the brachial plexus: Feasibility and initial experience in neoplastic conditions. *Neuroradiology* 2010;52:237-245.
32. Vargas MI, Gariani J, Delattre BA, Dietemann J-L, Lovblad K, Becker M. Three-Dimensional MR Imaging of the Brachial Plexus. Seminars in Musculoskeletal Radiology. 2015;19(2):137-48.

33. Tagliafico A, Calabrese M, Puntoni M, Pace D, Baio G, Neumaier CE, et al. Brachial plexus MR imaging: accuracy and reproducibility of DTI-derived measurements and fibre tractography at 3.0-T. *Eur Radiol*. 2011;21(8):1764-71.
34. Thiel W. Die Konservierung ganzer leichen in natu¨rlichen farben. *Ann Anat* 1992;174:185-195.
35. De Crop A, Bacher K, Van Hoof T, Smeets PV, Smet BS, Vergauwen M, Kiendys U, Duyck P, Verstraete K, D'Herde K, Thierens H. Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study. *Radiology* 2012;262:298-304.
36. De Crop A, Smeets P, Van Hoof T, Vergauwen M, Dewaele T, Van Borsel M, Achten E, Verstraete K, D'Herde K, Thierens H, Bacher K. Correlation of clinical and physical-technical image quality in chest CT: a human cadaver study applied on iterative reconstruction. *BMC Med Imaging* 2015 Aug 19;15(1):32.
37. Audenaert EA, Vigneron L, Van Hoof T, D'Herde K, van Maele G, Oosterlinck D, Pattyn C. In vitro validation and reliability study of electromagnetic skin sensors for evaluation of end range of motion positions of the hip. *Medical & Biological Engineering & Computing* 2011;49:1405-1412.
38. Van Hoof T, Gomes GT, Audenaert E, Verstraete K, Kerckaert I, D'Herde K. 3d computerized model for measuring strain and displacement of the brachial plexus following placement of reverse shoulder prosthesis. *Anatomical Record-Advances in Integrative Anatomy and Evolutionary Biology* 2008;291:1173-1185.
39. Chandarana H, Block TK, Rosenkrantz AB, Lim RP, Kim D, Mossa DJ, et al. Free-Breathing Radial 3D Fat-Suppressed T1-Weighted Gradient Echo Sequence A Viable Alternative for Contrast-Enhanced Liver Imaging in Patients Unable to Suspend Respiration. *Investigative Radiology*. 2011;46(10):648-53.
40. Zheng Z-Z, Shan H, Li X. Fat-Suppressed 3D T1-Weighted Gradient-Echo Imaging of the Cartilage With a Volumetric Interpolated Breath-Hold Examination. *American Journal of Roentgenology*. 2010;194(5):W414-W9.
41. Chang G, Deniz CM, Honig S, Egol K, Regatte RR, Zhu Y, et al. MRI of the Hip at 7T: Feasibility of Bone Microarchitecture, High-Resolution Cartilage, and Clinical Imaging. *J Magn Reson Imaging.* 2014;39(6):1384-93.
42. Fleiss JL, Levin BA, Paik MC (2003). Statistical methods for rates andproportions. 3rd ed. Hoboken, NJ: Wiley.
43. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.

44. Lawton CAF, Michalski J, El-Naqa I et al. Variation in the definition of clinical target volumes for pelvic nodal conformal radiation therapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2009;74:377-382.
45. Ost P, De Meerleer G, Vercauteren T et al. Delineation of the postprostatectomy prostate bed using computed tomography: interobserver variability following the EORTC delineation guidelines. *Int J Radiat Oncol Biol Phys* 2011;81:143-149.
46. Van de Velde J, Vercauteren T, De Gersem W, Wouters J, Vandecasteele K, Vuye P, et al. Reliability and accuracy assessment of radiation therapy oncology group-endorsed guidelines for brachial plexus contouring. *Strahlenther Onkol*. 2014;190(7):628-+.
47. Stapleford LJ, Lawson JD, Perkins C, Edelman S, Davis L, McDonald MW, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics*. 2010;77(3):959-66.
48. Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. *Radiat Onco.* 2014;9(1):173.
49. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2014, 41.
50. Rohlfing T, Russakoff DB, Maurer CR. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *Ieee Transactions on Medical Imaging*. 2004;23(8):983-94.
51. Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage*. 2006;33(1):115-26.
52. Svarer C, Madsen K, Hasselbalch SG, Pinborg LH, Haugbol S, Frokjaer VG, et al. MR-based automatic delineation of volumes of interest in human brain PET images using probability maps. *Neuroimage*. 2005;24(4):969-79.
53. Han X, Hibbard S, O'Connell N, et al. Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion. proc. *MICCAI Workshop Head & Neck Autosegmentation Challenge* 2010;297–304.
54. Warfield SK, Zou KH, Wells WM. Validation of image segmentation by estimating rater bias and variance. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*. 2008;366(1874):2361-75.
55. Coupe P, Manjon JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage*. 2011;54(2):940-54.

56. Veronesi U, Orecchia R, Maisonneuve P, Viale G, Rotmensz N, Sangalli C, et al. Intraoperative radiotherapy versus external radiotherapy for early breast cancer (ELIOT): a randomised controlled equivalence trial. *Lancet Oncology*. 2013;14(13):1269-77
57. Ahn S, Park SH, Lee KH. How to Demonstrate Similarity by Using Noninferiority and Equivalence Statistical Testing in Radiology Research. *Radiology*. 2013;267(2):328-38.
58. Van de Velde J, Audenaert E, Speleers B, Vercauteren T, Mulliez T, Vandemaele P, et al. An Anatomically Validated Brachial Plexus Contouring Method for Intensity Modulated Radiation Therapy Planning. *International Journal of Radiation Oncology Biology Physics*. 2013;87(4):802-8.
59. Min M, Roos D, Keating E, Penniment M, Carruthers S, Zanchetta L, et al. External evaluation of the Radiation Therapy Oncology Group brachial plexus contouring protocol: Several issues identified. *J Med Imag Radiat Oncol*. 2014;58(3):360-8.
60. Genovesi D, Perrotti F, Trignani M, Di Pilla A, Vinciguerra A, Augurio A, et al. Delineating brachial plexus, cochlea, pharyngeal constrictor muscles and optic chiasm in head and neck radiotherapy: a CT-based model atlas. Radiologia Medica. 2015;120(4):352-60.
61. Yang J, Amini A, Williamson R, Zhang L, Zhang Y, Komaki R, Liao Z, Cox J, Welsh J, Court L, Dong L. Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy. *Pract Radiat Oncol*. 2013;3(4):e139-47.
62. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage*. 2009;46(3):726-38.
63. Pirozzi, Horvat M, Piper J, et al. Atlas-based Segmentation: Evaluation of a Multi-Atlas Approach for Lung Cancer. *Med Phys*. 2012;39:3677.
64. Van Rikxoort EM, Isgum I, Arzhaeva Y, et al. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. *Medical Image Analysis* 2010;14**:**39-49.
65. Rohlfing T, Russakoff DB, Maurer CR. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *Ieee Transactions on Medical Imaging*. 2004;23(8):983-94.
66. Crum WR, Griffin LD, Hill DLG, Hawkes DJ. Zen and the art of medical image registration: correspondence, homology, and quality. *Neuroimage*. 2003;20(3):1425-37.
67. Langerak TR, van der Heide UA, Kotte ANTJ, Viergever MA, van Vulpen M, Pluim JPW. Label Fusion in Atlas-Based Segmentation Using a Selective and Iterative Method for Performance Level Estimation (SIMPLE). *Ieee Transactions on Medical Imaging*. 2010;29(12):2000-8.
68. Artaechevarria X, Munoz-Barrutia A, Ortiz-de-Solorzano C. Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. *Ieee Transactions on Medical Imaging*. 2009;28(8):1266-77.

69. Ramus L, Commowick O, Malandain G. Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces. In: Jiang T, Navab N, Pluim JPW, Viergever MA, editors. Medical Image Computing and Computer-Assisted Intervention - Miccai 2010, Pt Iii. Lecture Notes in Computer Science. 63632010. p. 155-62.
70. Langerak TR, Berendsen FF, Van der Heide UA, Kotte ANTJ, Pluim JPW. Multiatlas-based segmentation with preregistration atlas selection. *Medical Physics*. 2013;40(9).
71. Thomas, TO, Refaat T, Choi M, et al. Brachial plexus dose tolerance in head and neck cancer patients treated with sequential intensity modulated radiation therapy. *Radiation Oncology.* 2015;10.
72. Lu Q, Hua J, Kassir MM, Delproposto Z, Dai Y, Sun J, et al. Imaging Lymphatic System in Breast Cancer Patients with Magnetic Resonance Lymphangiography. *Plos One*. 2013;8(7).
73. Yi SK, Mak W, Yang CC, Liu T, Cui J, Chen AM, et al. Development of a Standardized Method for Contouring the Lumbosacral Plexus: A Preliminary Dosimetric Analysis of this Organ at Risk among 15 Patients Treated with IMRT for Lower Gastrointestinal Cancers and the Incidence of Radiation Induced Lumbosacral Plexopathy. *International Journal of Radiation Oncology Biology Physics.* 2011;81(2):S172-S3.
74. Simmat I, Georg P, Georg D, Birkfellner W, Goldner G, Stock M. Assessment of accuracy and efficiency of atlas-based autosegmentation for prostate radiotherapy in a variety of clinical conditions. Strahlenther Onkol. 2012;188:807-813.
75. La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. Radiat Oncol. 2012;7:160.

8. Appendices

8.1 List of publications

1. **A1, Q1, IF: 4.176**. **Van de Velde J**, Audenaert EA, Speleers B, Vercauteren T, Mulliez T, Vandemaele P, Achten E, Kerckaert I, D'Herde K, De Neve W, Van Hoof T. An anatomically validated brachial plexus contouring method for intensity modulated radiation therapy planning. *Int J Radiat Oncol Biol Phys* 2013;87:802-808. Published
2. **A1, Q1, IF: 2.914**. **Van de Velde J**, Vercauteren T, De Gersem W, Wouters J, Vandecasteele K, Vuye P, Vanpachtenbeke F, D'Herde K, Kerckaert I, De Neve W, Van Hoof T. Reliability and accuracy assessment of radiation therapy oncology group-endorsed guidelines for brachial plexus contouring. *Strahlenther Onkol* 2014;190:628-635. Published
3. **Reply letter to the editor, IF: 4.176.** **Van de Velde J**, De Neve W, Van Hoof T. An anatomically validated brachial plexus contouring method for intensity modulated radiation therapy planning: reply. *Int J Radiat Oncol Biol Phys* 2014;89:224-225. Published
4. **A1, Q1 IF: 4.176.** **Van de Velde J**, Wouters J, Vercauteren T, De Gersem W, Duprez F, De Neve W, Van Hoof T. Morphometric atlas selection for automatic brachial plexus segmentation. *Int J Radiat Oncol Biol Phys* 2015;92(3):691-8. Published
5. **A1, Q2, IF: 2.36.** **Van de Velde J**, Wouters J, Vercauteren T, De Gersem W, De Neve W, Van Hoof T. Optimal number of atlases and label fusion for multi-atlas based automatic brachial plexus contouring in IMRT treatment planning. Radiat Oncol *2015*. Published
6. **A1, Q3, IF: 1.33**. **Van de Velde J**, Bogaert S, Vandemaele P, Huysse W, Leijnse J, Achten E, De Neve W, Van Hoof T. Brachial plexus 3D reconstruction from MRI with dissection validation – a baseline study for clinical applications. *Surg Radiol Anat 2015.* Published
7. **A1, Q2, IF: 2.36.** **Van de Velde J**, Wouters J, Vercauteren T, De Gersem W, Achten E, De Neve W, Van Hoof T. The effect of morphometric atlas selection on multi-atlas based automatic brachial plexus segmentation. *Radiat Oncol 2015*. Published
8. **A1, Q1, IF: 4.363.** Mulliez T, **Van de Velde J**, Veldeman L, De Gersem W, Vercauteren T, Speleers B, Degen H Wouters J, Van Hoof T, Van Greveling A, Monten C, Berwouts D, De Neve W. Deep inspiration breath hold in the prone position retracts the heart from the breast and internal mammary lymph node region. *Radiother Oncol* 2015. Published.
9. **A1, Q2, IF: 2.36.** De Puysseleyr A, **Van De Velde J**, Speleers B, Vercauteren T, Goedgebeur A, Van Hoof T, Boterberg T, De Neve W, De Wagter C, Ost P. Hair-sparing whole brain radiotherapy with volumetric arc therapy in patients treated for brain metastases: dosimetric and clinical results of a phase II trial. *Radiat Oncol* 2014;29:9:170. Published
10. **A1, Q1, IF: 2.914.** De Puysseleyr A, Mulliez T, Gulyban A, Bogaert E, Vercauteren T, Van Hoof T, **Van de Velde J**, Van Den Broecke R, De Wagter C, De Neve W. Improved cone-beam

computed tomography in supine and prone breast radiotherapy. Surface reconstruction, radiation exposure, and clinical workflow. *Strahlenther Onkol* 2013;189:945-50. Published

8.2 Developed contouring guidelines for manual BP contouring

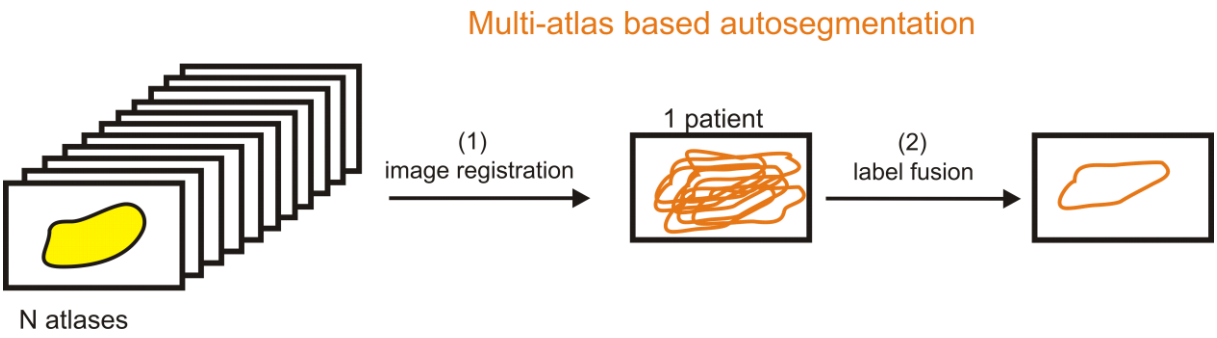
http://users.ugent.be/~jjvdevel/Delineation_guidelines.pptx

8.3 Terms and definitions

Multi-atlas-based automatic segmentation

Multi-atlas-based autosegmentation makes use of a database of multiple presegmented image datasets (called 'atlases') in order to achieve a new de novo BP delineation on a patient.

It consists of 2 major steps. An image registration and a label fusion.



1. Image registration

Image registration is the process of transforming different sets of data into one coordinate system.

The goal of image registration is to determine a common coordinate system in which images can be compared or analyzed.

Image transformation

The result of an image registration is an image transformation that realigns 2 or more planar images (2D) or image volumes (3D).

An image transformation is expressed as a mathematical function.

For rigid and affine transformations, this mathematical function is a transformation matrix.

For non-linear transformations, this mathematical function is a deformation vector field.

1. Rigid

Rigid transformation preserves distances between every pair of point. The rigid transformations include rotations, translations, reflections, or their combination.

Any object will keep the same shape and size after a proper rigid transformation.

2. Non-rigid or deformable

If a transformation changes the shape and/or size of a figure, then this transformation is non-rigid. Non-rigid transformations include affine transformations such as scaling and shear mapping but also typically involves non-linear transformation.

2.1 Affine

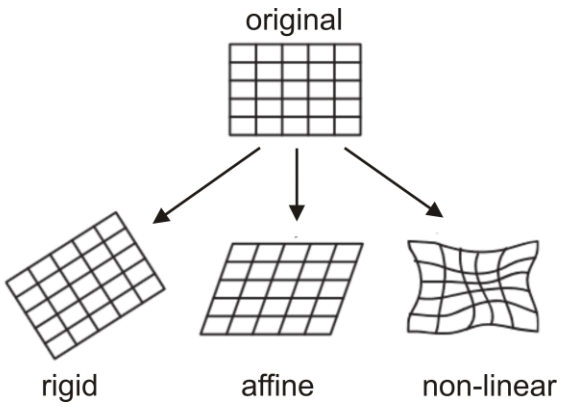
An affine transformation preserves points, straight lines and planes. Also, sets of parallel lines remain parallel after an affine transformation. An affine transformation does not necessarily preserve angles between lines or distances between points, though it does preserve ratios of distances between points lying on a straight line.

Examples of affine transformations include translation, scaling, reflection, rotation, shear mapping, and compositions of them in any combination and sequence.

All rigid transformations are examples of affine transformations.

2.2 Non-linear

A non-linear transformation can map straight lines into curves. It accommodates for local shape differences



By means of the 'transformation matrix' or 'deformation vector field', delineations of an organ on one image data sets can be deformed on a new image data set. This is what happens in the initial stage of the multi-atlas-based autosegmentation process.

2. Label fusion

In a second stage, the multiple deformed delineations on the new image data set are combined by the label fusion algorithm to obtain a unique and final consensus segmentation.

Label is a synonym for image segmentation.

2.1 STAPLE

Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm is an algorithm that was originally designed for the validation of image segmentations.

It takes a collection of segmentations and computes simultaneously a probabilistic estimate of the true segmentation and a measure of the performance level represented by each segmentation.

The STAPLE algorithm is more robust than simple averaging, as the latter is sensitive to outliers.

It ignores the image data but it only uses the segmentations when computing the label fusion.

2.2 Patch

Patch label fusion is an intensity-weighted label fusion algorithm. This means that for the calculation of the label fusion, accuracy of the initial deformable image registration is assessed by comparing the intensity (or gray contrast) similarity between the atlas image and the subject image, after deformable registration.

Therefore, the algorithm compares the similarity of intensity between corresponding 'patches' (small subvolumes of the image defined as 3D cubes) in the atlas and the subject image after registration. It achieves labeling of each voxel individually by comparing its surrounding patch with patches in the atlas image in which the labels of the central voxels are known. The similarity in intensity between the patch in the subject and the patch in the

atlas are used to perform a weighted label fusion. This weighting is performed to account for errors in the initial deformable image registration

Similarity coefficients

DSC

The DSC measures the spatial overlap between the gold standard A and the registered segmentation B and is defined as DSC (A,B)= 2(A∩B)/(A+B), where ∩ indicates the intersection between volume A and B. The DSC is situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement.

JI

The Jaccard similarity Index (JI) is the ratio of the intersection volume and the entire union volume (U) of the delineations JI(A,B)= (A∩B)/(AUB) . The JI is also situated between 0 and 1, with 0 indicating no agreement and 1 indicating perfect agreement. The JI can be calculated from DSC with the formula [JI = DSC/(2-DSC)].

INI or TPR

Finally, the Inclusion Index (INI) was measured. INI is the intersection volume between the gold standard (A) and the registered segmentation (B), divided by the gold standard INI=(A∩B/A). INI is situated between 0 and 1, with 0 indicating no inclusion and 1 indicating total inclusion of A by B.

The Jaccard index is a stricter index for the evaluation of contour overlap compared to Dice in the sense that using Jaccard index, the penalty for false positive delineation area is larger than using Dice. The major difference between Inclusion index (TPR) and the other two similarity indices is that no penalty is given for false positive delineation area using Inclusion index.

To illustrate the differences between the similarity indices, we studied their behavior with an increasing correctly delineated volume for 2 volumes with a 10 percent volume difference. In this example, the delineated volume was considered to be 10 percent larger than the gold standard volume, which is a realistic difference in volume with respect to the registered BPs. The results of this test are shown in figure 1. DSC and INI both increase linearly with the increase of overlap between the 2 segmentations, whereas JI has a non-linear increase. The reason for this difference is that with

JI, the penalty increases as the false positive delineation area increases. With INI in contrast, no penalty for false positive delineation area is given.

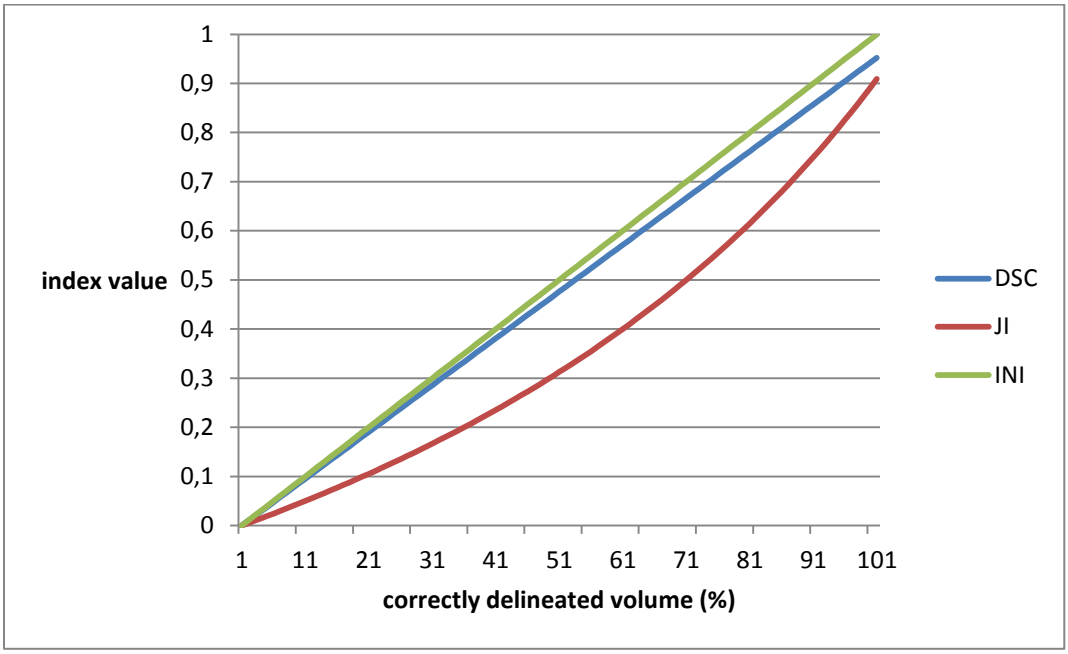


Fig. 1. Similarity indices as a function of correctly delineated volume

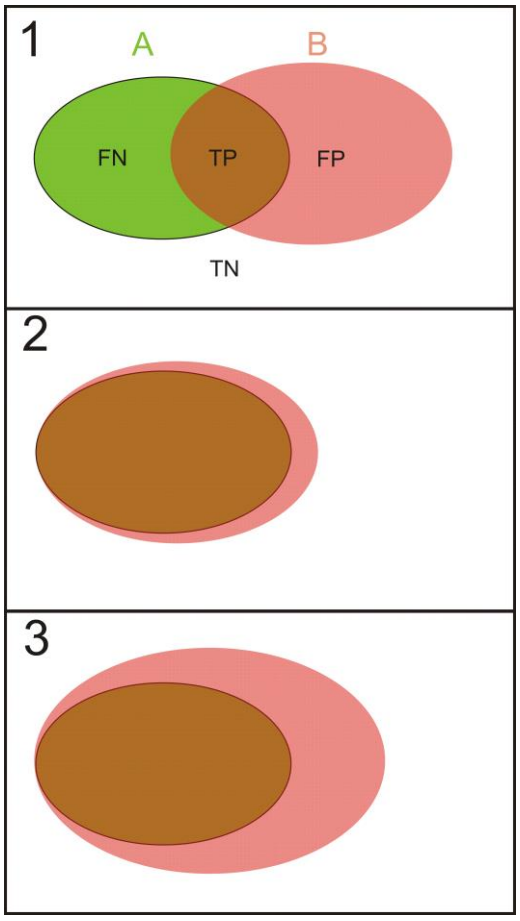


Fig.2 Illustration of 3 different delineation possibilities.

1, illustrates the gold standard segmentation A and the delineation B, that covers a part of the gold standard. TP, true positive delineation area; FP, false positive delineation area; FN, false negative delineation area; TN, true negative delineation area. 2, illustrates the entire inclusion of the gold standard A by the delineation B with a small false positive delineation area.3, illustrates the entire inclusion of the gold standard A by the delineation B with a large false positive delineation area

In Figure 2, three different delineation situations are illustrated. All three indices (DSC, JI and INI) increase with increasing true positive delineation area.

INI increases linearly up to 1 until the entire gold standard A is included (Fig 2.2). When the entire gold standard is included and the false positive delineation area increases, INI stays 1 (Fig 2.3).

DSC also increases linearly, because the denominator (A+B) stays the same value, regardless of the overlap. However, it penalizes false positive delineation area. When the delineation B is larger than

the gold standard A (Fig 2.2), the perfect value of 1 can never be reached. The larger the false positive delineation area, the smaller the DSC value (Fig 2.3).

JI increases non-linearly because the denominator AUB is changing depending on the overlap. JI penalizes false positive delineation area more than DSC, which makes that, with increasing true positive delineation area, JI has the lowest values of all three (Fig.1).

References

1. 1.Warfield SK, Zou KH, Wells WM. Validation of image segmentation by estimating rater bias and variance. In: Larsen R, Nielsen M, Sporring J, editors. Medical Image Computing and Computer-Assisted Intervention - Miccai 2006, Pt 2. Lecture Notes in Computer Science. 41912006. p. 839-47.
2. Warfield SK, Zou KH, Wells WM. Validation of image segmentation by estimating rater bias and variance. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences. 2008;366(1874):2361-75.
3. Teguh DN, Levendag PC, Voet PWJ, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. International Journal of Radiation Oncology Biology Physics. 2011;81(4):950-7.
4. Han X, Hoogeman MS, Levendag PC, Hibbard LS, Teguh DN, Voet P, et al. Atlas-Based Auto-segmentation of Head and Neck CT Images. In: Metaxas D, Axel L, Fichtinger G, Szekely G, editors. Medical Image Computing and Computer-Assisted Intervention - Miccai 2008, Pt Ii, Proceedings. Lecture Notes in Computer Science. 52422008. p. 434-41.
5. Coupe P, Manjon JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Neuroimage. 2011;54(2):940-54.
6. Bai W, Shi W, O'Regan DP, Tong T, Wang H, Jamil-Copley S, et al. A Probabilistic Patch-Based Label Fusion Model for Multi-Atlas Segmentation With Registration Refinement: Application to Cardiac MR Images. Ieee Transactions on Medical Imaging. 2013;32(7):1302-15.
7. Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. J Med Imag Radiat Oncol. 2010;54(5):401-10.

8. Simmat I, Georg P, Georg D, Birkfellner W, Goldner G, Stock M. Assessment of accuracy and efficiency of atlas-based autosegmentation for prostate radiotherapy in a variety of clinical conditions. Strahlenther Onkol. 2012;188(9):807-13.
9. Fotina I, Lutgendorf-Caucig C, Stock M, Potter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol. 2012;188(2):160-7.